

中英文对照

X-ray Crystallography
of Biomacromolecules

生物大分子的

X

射线晶体学

柯衡明 (Hengming Ke) 陈玉祥 蔡继文 编著



化学工业出版社

ISBN 7-03-014244-3

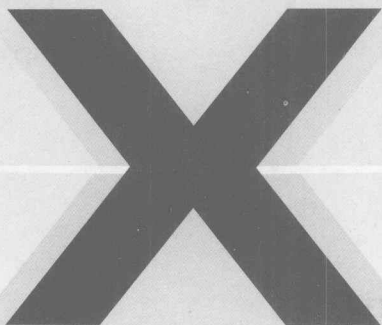
X-ray Crystallography
of Biomacromolecules
生物大分子的

射线晶体学

邢其毅 (Xing Qiyi) 编 王德明 (Wang Deming) 副编

X-ray Crystallography of Biomacromolecules

生物大分子的



射线晶体学

柯衡明 (Hengming Ke) 陈玉祥 蔡继文 编著



化学工业出版社

·北京·

本书主要介绍了生物大分子 X 射线衍射晶体学的基本原理, 包括蛋白质的纯化和晶体的培养、晶体的表征、X 射线衍射原理、蛋白质晶体结构分析的主要方法和手段、重原子衍生物的制备、分子置换法解析晶体结构、蛋白质模型的构建、电子密度图的解析以及结构精修等。本书结合第一作者柯衡明教授多年从事蛋白质晶体结构分析所积累的经验, 总结了在蛋白质纯化、蛋白质及其重原子衍生物晶体的培养、数据处理和结构分析过程中常见的问题和解决方法。

本书可作为从事结构生物学研究的科研工作者的参考书和研究生的教材。

图书在版编目 (CIP) 数据

生物大分子的 X 射线晶体学/柯衡明 (Ke, H.),
陈玉祥, 蔡继文编著. —北京: 化学工业出版社,
2010. 1

ISBN 978-7-122-06953-5

I. 生… II. ①柯…②陈…③蔡… III. 生物高
分子-X 射线晶体学 IV. ①Q71

中国版本图书馆 CIP 数据核字 (2009) 第 196234 号

责任编辑: 傅四周 黎秀芬

文字编辑: 周 侗

责任校对: 周梦华

装帧设计: 张 辉

出版发行: 化学工业出版社 (北京市东城区青年湖南街 13 号 邮政编码 100011)

印 装: 化学工业出版社印刷厂

720mm×1000mm 1/16 印张 12 字数 200 千字 2010 年 1 月北京第 1 版第 1 次印刷

购书咨询: 010-64518888(传真: 010-64519686) 售后服务: 010-64518899

网 址: <http://www.cip.com.cn>

凡购买本书, 如有缺损质量问题, 本社销售中心负责调换。

定 价: 39.00 元

版权所有 违者必究

前 言

人类基因组测序的基本完成，标志着生命科学进入后基因时代，其发展的必然结果之一就是“结构基因组学”的兴起。当前，揭示基因组功能的基本途径之一就是系统研究蛋白质的三维结构，即以基因组序列为基础进行系统的结构生物学研究。生物大分子，包括蛋白质、核酸和病毒等物质的三维空间结构是从分子水平上理解生命基本过程的重要依据，以其为主体的结构生物学是生命科学的重要组成部分。近十几年来，基于三维结构的分子药物设计的迅速发展使结构生物学成为分子生命科学的前沿之一。

作为结构生物学的主流研究手段，生物大分子 X 射线衍射技术已经广泛应用于生物、化学、医学和药学等领域。尤其是在过去的 10 年里，随着数据处理和结构分析软件的发展和完善、单晶衍射仪价格的降低和普及，晶体结构分析工作也由原来晶体学家的专业工作过渡到常规的分析手段，成为很多生物和化学工作者日常科研工作的一部分。本书以结构测定的流程为线索，结合作者多年的晶体学研究 and 应用经验，深入浅出地介绍晶体学基本原理及最新的晶体结构分析技术和方法。本书重点介绍蛋白质晶体培养方法、同晶置换、多波长反常衍射、相位精修、分子置换的基本原理，重原子衍生物制备的基本过程和方法，电子密度图解析，结构模型构建和最后的结构精修。因此，本书对结构生物学和分子药学工作者具有重要的参考价值。

柯衡明 (Hengming Ke) 博士 1989 年在哈佛大学获生物物理学博士学位，现为北卡罗莱纳大学医学院教授，美国结晶学协会、美国科学进步协会委员，中南大学、中山大学客座教授。柯衡明教授主要从事药物靶标蛋白和具有重要生理功能的蛋白酶的结构与功能、相关抑制剂的设计研究工作，在“Science”和“PNAS”等学术期刊发表学术论文和研究综述 60 余篇，担任本书的第一编著者，主要负责英文稿的写作。中南大学教授陈玉祥博士和中山大学原教授蔡继文博士 (女) 主要负责中文稿的编译和英文稿的修订。本书在编译过程中得到了中南大学湘雅医院杨菊云老师，中南大学生物科学与技术学院研究生周志群、吉午阳、郭婷婷、谢薇、李波、李一卉、张利元、颜世能、危丽雯、周艳琴、李辉莹，中山大学药学院研究生严子耳、罗璇、罗敏贤、刘明、徐洁以及老师罗海彬和陈芸芸的帮助，在此一并致谢。

目 录

1 大分子结构测定概述	1
参考文献	7
2 生物大分子的结晶	8
2.1 结晶的理论基础	8
2.2 影响结晶的因素	10
2.3 结晶方法	17
2.4 结晶的实用心得	20
2.5 结晶过程的自动化	27
参考文献	28
3 晶体的表征	29
3.1 晶系	29
3.2 对称性与空间群	33
3.3 对称性的矩阵表示	38
3.4 晶胞的不对称单位	39
3.5 倒易晶胞	41
参考文献	43
4 X 射线衍射原理	44
4.1 X 射线的产生	44
4.2 一个电子的散射	46
4.3 一个原子对 X 射线的散射	48
4.4 一个分子对 X 射线的散射	50
4.5 衍射的中心对称性	53
4.6 布拉格衍射定律	54
4.7 爱瓦尔德衍射球	55
4.8 衍射空间的对称性	58
5 帕特森函数	63
5.1 相角问题简介	63
5.2 帕特森函数	64
5.2.1 帕特森函数的理论基础	64
5.2.2 帕特森图的解析	68
参考文献	74

6	同晶置换	75
6.1	同晶置换理论	75
6.2	相角计算	77
6.3	重原子的最小二乘法精修	79
6.4	重原子最大概率精修	81
	参考文献	83
7	多波长反常衍射	84
7.1	多波长反常衍射的原理	84
7.2	反常散射的差异	88
7.3	吸收限	89
7.4	MAD 的实验设计	91
7.5	MAD 中相位的计算	93
	参考文献	96
8	相位优化	97
8.1	溶剂平滑法	97
8.2	直方图匹配	100
8.3	局部或非晶体学对称性平均	102
8.4	相位组合	103
	参考文献	104
9	重原子衍生物的制备	105
9.1	同晶现象	105
9.2	溶液中影响重金属结合的原因	108
9.2.1	pH 值的影响	108
9.2.2	缓冲液	109
9.2.3	沉淀剂	110
9.2.4	重金属浓度、浸泡时间和温度	111
9.3	重金属概论	112
9.3.1	汞	113
9.3.2	铂、金、钯	115
9.3.3	铀和钷	117
9.3.4	铀酸盐和阳离子试剂	118
9.3.5	氙	120
9.4	MAD 实验用的重原子	120
9.5	硒代甲硫氨酰蛋白质突变体的制备	121

9.6 硒代甲硫氨酸内切酶 <i>Nae</i> I 表达的实验设计 (Ke's Lab, UNC, August 1998)	123
9.6.1 大肠杆菌细胞的增殖	123
9.6.2 改良的 M9 培养基的制备	124
参考文献	125
10 分子置换法	126
10.1 引言	126
10.2 坐标转换	128
10.3 欧拉角和极化角	130
10.4 分子置换法原理	132
10.5 旋转函数	133
10.6 平移函数	136
10.7 分子对称性和自旋转函数	140
参考文献	142
11 蛋白质的结构	144
11.1 多肽的构象	144
11.2 二级结构	149
参考文献	152
12 电子密度图的解析与模型构建	153
12.1 电子密度图	153
12.2 分辨率与图的质量	154
13 结构精修	161
13.1 判断结构的正确性	161
13.2 最小二乘法	163
13.3 立体化学限制的最小二乘法精修	168
13.4 分子动力学精修法	169
13.4.1 鲍威尔能量最小化	169
13.4.2 分子动力学	173
参考文献	175
常见单词翻译	176

Content

1 Overview of Structure Determination of Macromolecules	1
References	7
2 Crystallization of Biomacromolecules	8
2.1 Theoretical basis of crystallization	8
2.2 Factors affecting crystallization	10
2.3 Methods of crystallization	17
2.4 Practical aspects of crystallization	20
2.5 Automation of crystallization	27
References	28
3 Characterization of Crystals	29
3.1 Crystal systems	29
3.2 Symmetry and space group	33
3.3 Matrix representation of symmetry	38
3.4 Asymmetric unit of unit cell	39
3.5 Reciprocal unit cell	41
References	43
4 Principle of X-ray Diffraction	44
4.1 X-ray source	44
4.2 X-ray scattering by an electron	46
4.3 X-ray scattering by an atom	48
4.4 X-ray scattering by a molecule	50
4.5 Center symmetry of reflection intensity	53
4.6 Bragg's law of diffraction	54
4.7 Ewald reflection sphere	55
4.8 Symmetry in diffraction space	58
5 Patterson Function	63
5.1 Introduction to phase problem	63
5.2 Patterson function	64
5.2.1 Theory of Patterson function	64
5.2.2 Interpretation of Patterson map	68
References	74

6	Isomorphous Replacement	75
6.1	Principle of isomorphous replacement	75
6.2	Phase calculation	77
6.3	Least-square heavy atom refinement	79
6.4	Maximum-likelihood refinement of heavy atom	81
	References	83
7	Multiwavelength Anomalous Diffraction	84
7.1	Theory of multiwavelength anomalous diffraction	84
7.2	Anomalous scattering difference	88
7.3	Absorption edge	89
7.4	Experimental design of MAD	91
7.5	Phase calculation of MAD	93
	References	96
8	Phase Improvement	97
8.1	Solvent flattening	97
8.2	Histogram matching	100
8.3	Local or non-crystallographic symmetry averaging	102
8.4	Phase combination	103
	References	104
9	Preparation of Heavy Atom Derivatives	105
9.1	Isomorphism	105
9.2	Factors affecting heavy metal binding to crystals	108
9.2.1	Effect of pH	108
9.2.2	Buffer	109
9.2.3	Precipitant	110
9.2.4	Heavy metal concentration, soaking time and temperature	111
9.3	Survey of heavy metals	112
9.3.1	Mercury	113
9.3.2	Platinum, gold, and palladium	115
9.3.3	Iridium and Osmium	117
9.3.4	Uranate and cation reagents	118
9.3.5	Xenon	120
9.4	Heavy atom for MAD	120
9.5	Preparation of selenomethionyl mutant of proteins	121
9.6	Protocol for expression of selenomethionyl endonuclease <i>Nae</i> I (Ke's Lab, UNC, August 1998)	123
9.6.1	Growth of <i>E. coli</i> cell	123
9.6.2	Preparation of the modified M9 medium	124
	References	125

10	Molecular Replacement	126
10.1	Introduction	126
10.2	Coordinate transformation	128
10.3	Eulerian and Polar angles	130
10.4	Principle of molecular replacement	132
10.5	Rotation function	133
10.6	Translation function	136
10.7	Molecular symmetry and self-rotation function	140
	References	142
11	Architecture of Protein Structures	144
11.1	Conformation of a polypeptide	144
11.2	The secondary structure	149
	References	152
12	Electron Density Interpretation and Model Building	153
12.1	Electron density map	153
12.2	Resolution and map quality	154
13	Structure Refinement	161
13.1	Judgment on correctness of a structure	161
13.2	Least-square refinement	163
13.3	Stereochemically restrained least-square refinement	168
13.4	Molecular dynamic refinement	169
13.4.1	Powell energy minimization	169
13.4.2	Molecular dynamics	173
	References	175

1

大分子结构测定概述

Overview of Structure Determination of Macromolecules

大分子晶体结构测定可以分为以下七个步骤。

第一步：结晶。利用 X 射线衍射法进行三维结构的分析，首先必须获得合适的单晶。获得生物大分子晶体的方法主要有蒸汽扩散法和透析结晶法。晶体的衍射能力要高于 $3\text{\AA}^{\text{①}}$ ，才能确定结构中原子的位置（图 1.1）。

第二步：晶体的初步表征。对晶体进行初步的衍射实验（图 1.2）有两个目的：①确定晶系、晶胞参数和空间群；②检查晶体质量与衍射能力。

第三步：衍射数据收集。目前，衍射数据的收集已经是自动化的流程（图 1.3）。最常用于收集数据的 X 射线面探测器是影像板成像探测器（IP）和电荷耦合探测器（CCD）。Multiwire 探测器和 TV 探测器被人们摒弃了，照相法早已经成为历史。

Structure determination of macromolecular crystals can be divided into seven steps.

Step 1. Crystallization. Single crystals are required for the analysis of three-dimensional structures by X-ray diffraction. Vapor diffusion and dialysis are two major methods to obtain single crystals of biomacromolecules. A crystal diffracting to $3\text{\AA}^{\text{①}}$ or better resolution is required for determination of a structure with resolved positions of individual atoms (Fig. 1.1).

Step 2. Preliminary characterization of crystals. The preliminary diffraction of crystals (Fig. 1.2) has two purposes: ① determination of the crystal system, unit cell parameters, and the space group, and ② examination of quality and diffraction capability of crystals.

Step 3. Data collection. Collection of diffraction data is an automatic procedure (Fig. 1.3). The most commonly used X-ray area detectors for data collection are phosphate image plate and charge coupled device (CCD). Multiwire and TV detectors are out of date. The film method becomes a history.

① $1\text{\AA}=0.1\text{nm}$ 。

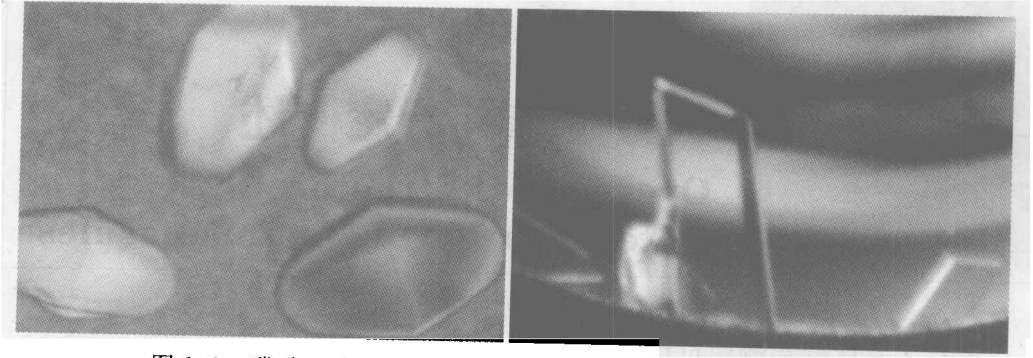


图 1.1 磷酸二酯酶 4B (PDE 4B, 左) 和 p21 (细胞周期蛋白
依赖激酶 2 的抑制剂, 右) 的晶体

PDE 4B 催化结构域的六角形晶体 (残基 152~505) 是用 16mg/mL PDE 蛋白在 4°C 对以下缓冲液进行透析得到: 20mmol/L 4-羟乙基哌嗪乙磺酸, pH7.0、50mmol/L 二甲基胍酸盐、0.2mol/L 醋酸钠、3%叔丁醇、5.5% PEG 3350。p21 的单斜晶体是对以下缓冲液透析获得: 20mmol/L Tris, 碱, pH8.5、0.5mmol/L NaN₃、0.5mmol/L 巯基乙醇、0.5mmol/L EDTA、6%PEG 3350

Fig. 1.1 Crystals of phosphodiesterase (left) and p21, an inhibitor of cyclin-dependent kinase 2 (right)

The hexagonal crystals of the catalytic domain of PDE 4 B (residues 152-505) were grown by dialyzing 16mg/mL PDE protein against a buffer of 20mmol/L HEPES, 50mmol/L cocadylate, 0.2mol/L sodium acetate, 3% *t*-butanol, 5.5% PEG3350 at pH7.0 and 4°C. The monoclinic crystals of p21 were grown by dialysis against 20mmol/L Tris, base, 0.5mmol/L NaN₃, 0.5mmol/L β -mercaptoethanol, 0.5mmol/L EDTA, 6% polyethylene glycol 3350 at pH 8.5

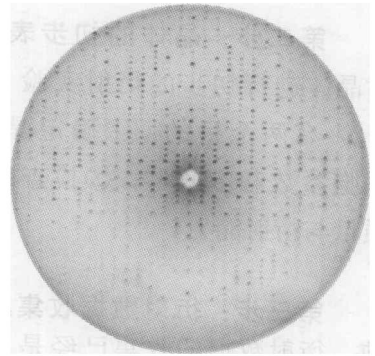
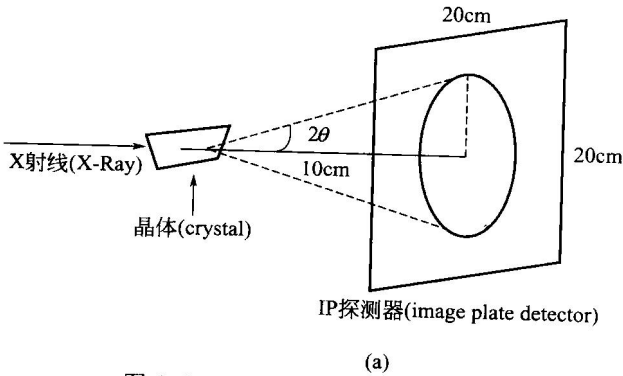


图 1.2 (a) 蛋白质晶体衍射示意; (b) 蛋白质晶体衍射的旋进照片
沿着垂直轴和水平轴方向的 2 重对称性清晰可见。

衍射点之间的间隔决定了晶胞参数
 a 、 b 、 c 、 α 、 β 和 γ 的大小

Fig. 1.2 (a) Schematic diagram of crystal diffraction;
(b) A precession photo of protein crystal diffraction

The 2-fold symmetry along the vertical and horizontal axes is visible.
The spacing and angles between the diffraction spots determine
the unit cell parameters a , b , c , α , β and γ

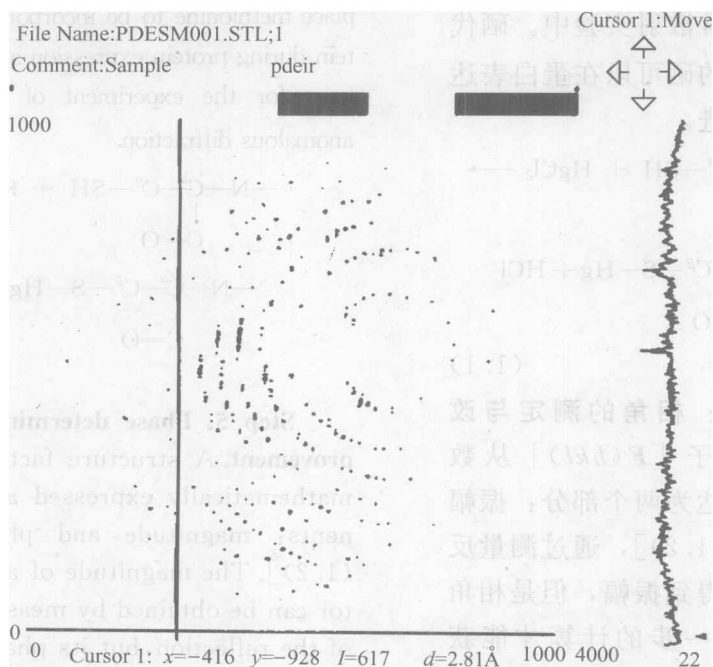


图 1.3 用 Rigaku IP 获得的一幅典型的衍射图像
每个衍射点的黑度代表衍射强度

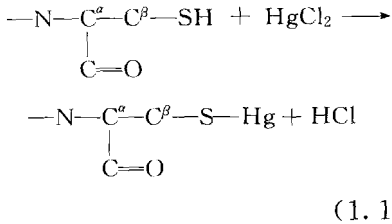
Fig. 1.3 A typical pattern of diffraction, as recorded in the
Rigaku image plate detector

The darkness of each spot represents its diffraction intensity

第四步：制备重原子衍生物。重原子衍生物对于大分子的结构测定是至关重要的。所谓重原子衍生物就是结合了重金属化合物的蛋白质晶体，如下面所示的汞共价结合半胱氨酸。它们可以通过将天然晶体浸泡在含有重金属化合物的缓冲液中，或将天然蛋白与重金属化合物一起共结晶。另一类重原子衍生物是硒代

Step 4. Preparation of heavy atom derivatives. Heavy atom derivatives are essential for structure determination of macromolecules. Protein crystals with binding of heavy metal compounds are called heavy atom derivatives. An example is the covalent binding of mercury to cysteine [equation (1.1)]. They can be prepared by soaking native crystals in a buffer containing a heavy metal compound or by co-crystallizing a protein with a heavy metal compound. Another category of heavy atom derivatives is the selenomethionyl mutants of proteins.

甲硫氨酰蛋白质的突变体，常用于多波长反常散射实验中。硒代甲硫氨酰中的硒可以在蛋白表达的过程中引进。

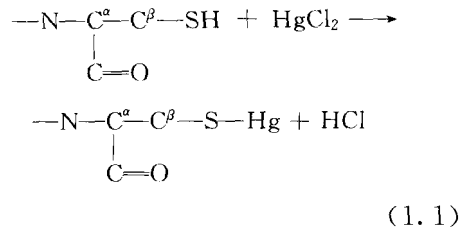


第五步：相角的测定与改进。结构因子 $[F(hkl)]$ 从数学上可以表达为两个部分：振幅和相角 [式(1.2)]，通过测量反射强度可以得到振幅，但是相角信息需要进一步的计算才能获得。相角确定是晶体结构分析中的核心问题。目前主要有五种方法用于确定大分子晶体结构的相角：①帕特森函数；②多重同晶置换 (MIR)；③多波长反常衍射 (MAD)；④分子置换法 (MR)；⑤直接法和最大熵法。

$$\begin{aligned} F(hkl) &= \\ \sum_j f_j \exp[2\pi i(hx_j + ky_j + lz_j)] & \\ = |F(hkl)| \exp(i\alpha) & \quad (1.2) \end{aligned}$$

获得初步的相位信息后，对相位信息进行改进是结构测定的另一步。一些常用的方法包括溶剂平滑法、直方图匹配法、局部或非晶体学对称性平均法、相位组合法、直接法和最大熵法。

Selenium in the selenomethionyl form can replace methionine to be incorporated into a protein during protein expression and is a key scatterer for the experiment of multiwavelength anomalous diffraction.



Step 5. Phase determination and improvement. A structure factor $[F(hkl)]$ is mathematically expressed as two components: magnitude and phase [equation (1.2)]. The magnitude of a structure factor can be obtained by measuring intensity of the reflection, but its phase needs to be calculated. Phase is the central problem in structure determination. There are mainly five methods for phasing macromolecular crystals: ① Patterson function, ② multiple isomorphous replacement (MIR), ③ multiwavelength anomalous diffraction (MAD), ④ molecular replacement (MR), ⑤ direct method and maximum entropy.

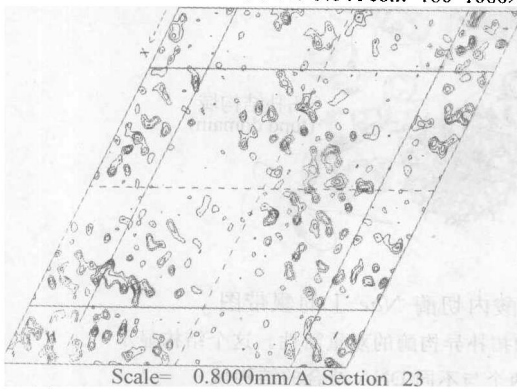
$$\begin{aligned} F(hkl) &= \sum_j f_j \exp[2\pi i(hx_j + ky_j + lz_j)] \\ &= |F(hkl)| \exp(i\alpha) \quad (1.2) \end{aligned}$$

After preliminary phases have been obtained, phase improvement is often the next step for structure determination. The commonly used methods of phase improvement include solvent flattening, histogram matching, non-crystallographic averaging, phase combination, direct methods and maximum entropy.

第六步：解析电子密度图与构建模型。利用实验测得的振幅和计算得到的相角进行逆傅里叶变换可以算出电子密度图 [式(1.3)]。电子密度图的质量取决于晶体衍射分辨率(图 1.4)。如今即使人们可以采用一些程序自动化地对电子密度图进行解析和模型构建,但由于电子密度图只显示峰的强度而不是原子模型,因此这个步骤在很大程度上取决于操作者的经验。有时,一个不好的电子密度图可能会产生错误的分子模型。

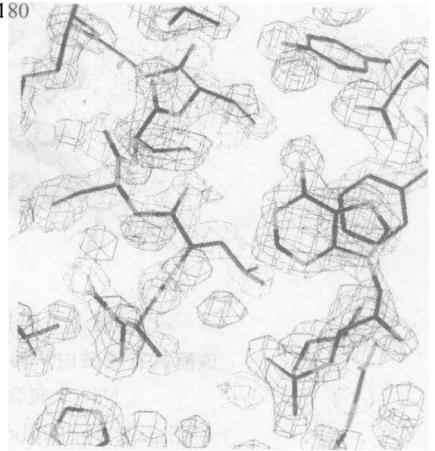
Step 6. Interpretation of electron density map and model building. An electron density map can be calculated by reverse Fourier transformation of structure factors that comprise the experimentally measured magnitudes and calculated phases [equation(1.3)]. The map quality is dependent on the resolution that a crystal can diffract to (Fig. 1.4). Although there are some programs available for automatic processing, interpretation of electron density map and model building still heavily depend on personal experience. A wrong model may be occasionally obtained from a marginally good electron density map.

FBPASE AVERAGING-FILTERING MAP 3.0 Å cont=180-1080×180



FBPase AVERAGING-FILTERING MAP P3221 3.0

(a)



(b)

图 1.4 (a) 在分辨率为 3\AA 时, 果糖-1,6-二磷酸酶电子密度的二维图; (b) 在分辨率为 1.6\AA , 等高线 1σ 时, 计算机的图形终端显示出的磷酸二酯酶 10A 电子密度图

位于图中左侧的螺旋和右下角的 cAMP 轮廓清晰可见

Fig. 1.4 (a) A two-dimensional plot of electron density map of fructose-1,6,-bisphosphatase at 3\AA resolution; (b) A portion of electron density of phosphodiesterase 10A that was calculated from the structure at 1.6\AA resolution and contoured at 1σ , as displayed in a graphic terminal

The helix on the left of the map and cAMP on the lower right are shaped very well

$$\rho(xyz) = (1/V) \sum_{hkl} |F(hkl)| \exp(i\alpha) \exp[-2\pi i(hx + ky + lz)] \quad (1.3)$$

第七步：结构精修。根据电子密度图构建的初始结构模型可以通过 CNS 程序中的最小二乘法或分子动力学法进行精修。一个正确的大分子结构，其 R 因子大约为 20%，并且具有较好的几何分子构型（图 1.5）。

$$\rho(xyz) = (1/V) \sum_{hkl} |F(hkl)| \exp(i\alpha) \exp[-2\pi i(hx + ky + lz)] \quad (1.3)$$

Step 7. Structure refinement. A raw structure model built from an electron density map is refined by a protocol of least-square or molecular dynamics, such as that implemented in program CNS. A correct structure of a macromolecule has an R-factor typically about 20% and also good molecular geometry (Fig. 1.5).

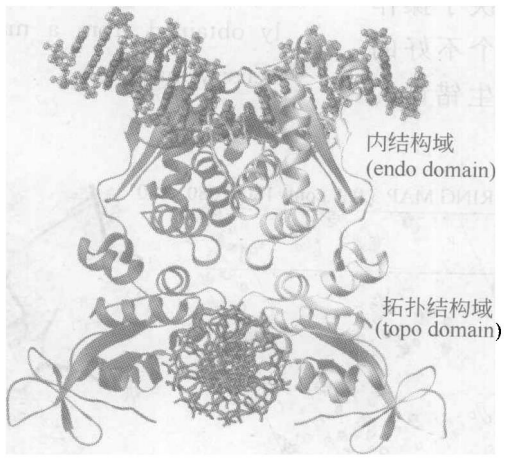


图 1.5 核酸内切酶 *Nae I* 的飘带图
该酶具有核酸内切酶和拓扑异构酶的双重活性。这个结构显示 *Nae I* 包含两个与不同 DNA 结合的结构域

Fig. 1.5 Ribbon diagram of endonuclease *Nae I*

It has dual activities of endonuclease and topoisomerase. The structure shows that *Nae I* contains two domains respectively responding for binding of the DNA

在这个教程中，着重介绍 X 射线晶体学的原理和基本概念，对于如何进行实际操作，并完成对大分子结构的测定，则需要要在晶体学实验室中学习。

In this introductory book, we will concentrate on understanding the principles and the basic concepts of X-ray crystallography. For the practical determination of macromolecular structures, you are encouraged to learn in a crystallographic laboratory.