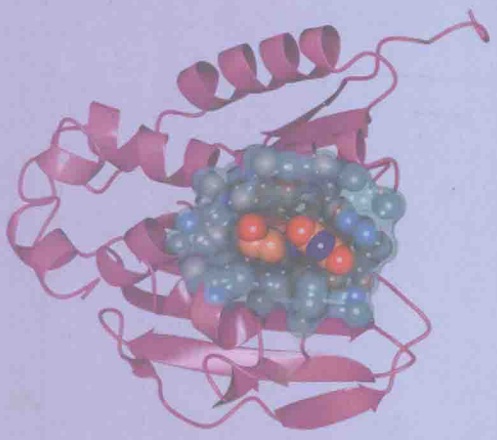
Methods in Molecular Biology 841

Springer Protocols



Leslie W. Tari *Editor*

Structure-Based Drug Discovery

Humana Press

Structure-Based Drug Discovery

Edited by

Leslie W. Tari

Trius Therapeutics, San Diego, CA, USA

常州大学山书馆
藏书章

Humana Press

Editor
Leslie W. Tari
Trius Therapeutics
San Diego, CA, USA
ltari@triusrx.com

ISSN 1064-3745 e-ISSN 1940-6029
ISBN 978-1-61779-519-0 e-ISBN 978-1-61779-520-6
DOI 10.1007/978-1-61779-520-6
Springer New York Dordrecht Heidelberg London

Library of Congress Control Number: 2011944430

© Springer Science+Business Media, LLC 2012
All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher (Humana Press, c/o Springer Science+Business Media, LLC, 233 Spring Street, New York, NY 10013, USA), except for brief excerpts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed is forbidden.
The use in this publication of trade names, trademarks, service marks, and similar terms, even if they are not identified as such, is not to be taken as an expression of opinion as to whether or not they are subject to proprietary rights.

Printed on acid-free paper

Humana Press is part of Springer Science+Business Media (www.springer.com)

METHODS IN MOLECULAR BIOLOGY™

Series Editor
John M. Walker
School of Life Sciences
University of Hertfordshire
Hatfield, Hertfordshire, AL10 9AB, UK

For further volumes:
http://www.springer.com/series/7651

Preface

The potential utility of atomic resolution structures of protein drug targets in drug discovery has long been acknowledged. Without structure, medicinal chemists must rely on the costly, time-consuming endeavor of screening large libraries of compounds for hits, and are often forced to live with high molecular weight, non-ligand-efficient inhibitor scaffolds that must be blindly decorated with thousands of groups to generate SAR, improve potency and properties. With knowledge of the shape and chemical composition of the ligand-binding pocket of the drug target, the de novo design of ligand efficient inhibitor scaffolds is enabled. Also, iterative-structure-guided ligand optimization can be used to rationally improve early leads in a few steps rather than with thousands of analogs. However, despite its promise, structure-based drug design (SBDD) did not live up to expectations in its early days: only a limited range of protein targets were tractable to crystallographic studies, crystal structures took months or years to solve, and limitations in computing power and unrealistic expectations of the capabilities of molecular modeling methods reduced the scope and effectiveness of SBDD.

The last decade has seen the confluence of several enabling technologies that have allowed protein crystallographic methods to live up to their true potential. Off-the-shelf systems exist that allow the rapid cloning, and recombinant expression and isolation of large quantities of protein in a wide range of prokaryotic or eukaryotic hosts. Low-cost nanovolume liquid-handling robotic systems are available for the automated screening of vast arrays of diverse solution conditions to find crystallization conditions for a protein target using minimal quantities of protein. Latest generation synchrotron radiation sources allow for the collection of high-resolution X-ray diffraction data on microcrystals in minutes. Continuing improvements in computing power and advances in crystallographic software have made it possible to go from X-ray dataset to refined crystal structure in less than an hour on a laptop computer. Taken together, these advances have made it possible to tackle difficult biological targets with a high probability of success: intact bacterial ribosomes have been structurally elucidated, as well as eukaryotic trans-membrane proteins like the potassium channel and GPCRs. Of additional importance is the impact the above mentioned advances have had on the throughput of crystallographic structure determinations: it is now possible for medicinal chemists to have access to structural information on their latest small molecule candidates bound to the therapeutic target within days of compound synthesis, allowing structure-guided ligand optimization to occur in "real time." Also, using fragment screening, crystal structures of hundreds of small molecule cores complexed with the protein target can be utilized to construct novel inhibitor scaffolds.

The goal of this book is to provide scientists interested in adding SBDD to their arsenal of drug discovery methods with a practical guide to the methods used to generate crystal structures of biological macromolecules, how to leverage the structural information to design new inhibitor classes de novo, and how to iteratively optimize hits and convert them to leads. Where possible, specific protocols are described. Some examples highlighting the utility of structural biology in the discovery and development of small molecule and protein therapeutic agents are provided in the later chapters.

此为试读,需要完整PDF请访问: www.ertongbook.com

I am deeply grateful to all contributors who agreed to share their experiences in the development and application of methodologies that support SBDD. I believe their patience and hard work will be rewarded by the impact this volume has on scientists involved in drug discovery. I would like to extend special thanks to John Walker for his guidance, inspiration and patience in the preparation of this volume. Also, I am grateful to Les Tari Sr. for his critical evaluation of this volume and sharp editorial eye.

San Diego, CA, USA *Leslie W. Tari*

Contributors

JUAN CARLOS ALMAGRO • *Centocor R&D Inc., Radnor, PA, USA*
JOHN BADGER • *Zenobia Therapeutics, San Diego, CA, USA*
DANIEL C. BENSEN • *Trius Therapeutics, San Diego, CA, USA*
MARK L. CUNNINGHAM • *Trius Therapeutics, San Diego, CA, USA*
JUDIT É. DEBRECZENI • *Structure and Biophysics, Discovery Sciences, AstraZeneca, Alderley Park, Macclesfield, UK*
SUZANNE C. EDAVETTAL • *Centocor R&D Inc., San Diego, CA, USA*
PAUL EMSLEY • *Department of Biochemistry, University of Oxford, Oxford, UK*
ANDREW D. FERGUSON • *Discovery Sciences, AstraZeneca Pharmaceuticals, Waltham, MA, USA*
JOHN FINN • *Trius Therapeutics, San Diego, CA, USA*
GARY L. GILLILAND • *Centocor R&D Inc., Radnor, PA, USA*
ISAAC D. HOFFMAN • *Takeda San Diego, San Diego, CA, USA*
MICHAEL J. HUNTER • *Centocor R&D Inc., San Diego, CA, USA*
ANDY JENNINGS • *Takeda San Diego, San Diego, CA, USA*
FELICE C. LIGHTSTONE • *Lawrence Livermore National Laboratory, Physical and Life Sciences Directorate, Livermore, CA, USA*
JINQUAN LUO • *Centocor R&D Inc., Radnor, PA, USA*
SHAWN P. MADDAFORD • *NeurAxonInc, Mississauga, ON, Canada L5K 1B3*
TOAN B. NGUYEN • *Lawrence Livermore National Laboratory, Physical and Life Sciences Directorate, Livermore, CA, USA*
GYORGY SNELL • *Takeda San Diego, San Diego, CA, USA*
RONALD V. SWANSON • *Centocor R&D Inc., San Diego, CA, USA*
LESLIE W. TARI • *Trius Therapeutics, San Diego, CA, USA*
OMID VAFA • *Centocor R&D Inc., Radnor, PA, USA*
SERGIO E. WONG • *Lawrence Livermore National Laboratory, Physical and Life Sciences Directorate, Livermore, CA, USA*

Contents

Chapter 1

The Utility of Structural Biology in Drug Discovery

Leslie W. Tari

Abstract

Access to detailed three-dimensional structural information on protein drug targets can streamline many aspects of drug discovery, from target selection and target product profile determination, to the discovery of novel molecular scaffolds that form the basis of potential drugs, to lead optimization. The information content of X-ray crystal structures, as well as the utility of structural methods in supporting the different phases of the drug discovery process, are described in this chapter.

Key words: X-ray crystallography, Structure-based drug design, Fragment screening, Structural bioinformatics, Lead optimization

1. Introduction

The discovery of new drugs is a time and labor-intensive process. On average, the discovery of a new drug requires the preparation and evaluation of approximately 10,000 compounds over 12 years at a cost of more than $350 million (1). Once in the marketplace, many drugs fail to recover their development costs (as many as 30%, according to data from the 1980s (2)), and many others are ultimately withdrawn from the market. These facts coupled with limits on patent lifetime, escalating global competition, and increasingly stringent government regulations for drug approval have demanded more efficient and accelerated approaches to drug discovery. Conventional "brute force" methods of lead discovery via high-throughput screening (HTS) of proprietary synthetic, combinatorial, or natural product libraries, while effective in many cases, are expensive and have limitations; they require access to large compound libraries (sometimes over 1,000,000 compounds), often yield hits with high molecular weight, poor ligand efficiency,

Leslie W. Tari (ed.), *Structure-Based Drug Discovery*, Methods in Molecular Biology, vol. 841,
DOI 10.1007/978-1-61779-520-6_1, © Springer Science+Business Media, LLC 2012

limited or no potential for optimization, and provide no information to guide ligand optimization.

Advances in crystallographic methods, computational power, molecular biology, and recombinant protein expression systems over the last 30 years have provided researchers with rapid and reliable access to three-dimensional structural information on a wide variety of protein drug targets. Structural information on protein–ligand complexes can eliminate much of the complexity involved in the discovery and optimization of prospective drug leads. Indeed, structure-guided drug design efforts have led to the discovery of high profile drugs in multiple therapeutic areas, including the peptidomimetic HIV protease inhibitors for the treatment of HIV, the neuraminidase inhibitor Tamiflu™ for the treatment of influenza, the carbonic anhydrase inhibitor dorzolamide for the treatment of glaucoma, and the thrombin inhibitor ximelagatran, an oral anticoagulant (3). Access to structural information on the target of interest can streamline all aspects of drug discovery, from target selection to lead discovery and optimization, using methods that are summarized in this chapter.

2. The Information Content of Protein Crystal Structures

Protein crystals, like any crystalline substance, are regular, three dimensionally periodic arrays of identical molecules or molecular complexes (see Fig. 1). A common misconception regarding protein crystal structures is that they are not representative of the protein in solution due to the influence of extensive intermolecular interactions present in the crystalline state. The idea that protein crystal structures are heavily biased by "solid state" artifacts arises from inaccurate comparisons made between protein crystals and crystals of small molecular weight compounds. Crystals of small molecules and proteins differ in ways that extend beyond the properties of their component molecules. Small-molecule crystals typically only comprise the small molecule, while protein crystals contain 25–90% solvent by volume, depending on the protein. The remaining volume in protein crystals is occupied by protein molecules, and is analogous to an ordered gel with large interstitial spaces between protein molecules. By comparison, the number of contacts made in relation to the molecular mass of the protein in protein crystals is smaller by orders of magnitude than it is for small-molecule crystals. This causes the mechanical stability and integrity of protein crystals to be much worse than it is for crystals of small molecules. The high solvent content and tenuous thermodynamic stability of protein crystals complicate the subsequent steps in X-ray diffraction experiments, since these properties result in crystal handling difficulties, susceptibility to temperature changes

此为试读,需要完整PDF请访问: www.ertongbook.com

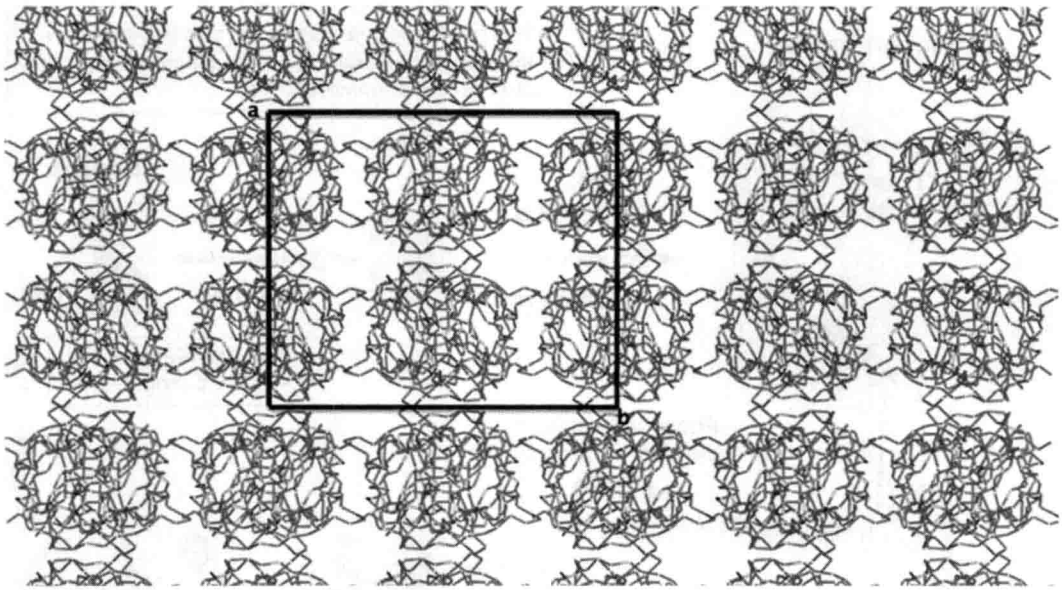


Fig. 1. A view of crystal packing in a *Haemophilus influenzae* dihydrofolate reductase crystal. Boundaries for a single unit cell within the crystal are shown. The view is perpendicular to the *c*-axis of the unit cell. The unit cell is the fundamental building block of the crystal, a translationally periodic substance comprising trillions of unit cells that extend in three dimensions. The unit cell is an arbitrary construction that describes the smallest "box" with the highest metric symmetry.

and dehydration, weaker diffraction, and greater sensitivity to radiation damage. However, the key role played by solvent in protein crystallization is a double-edged sword; while it adversely affects diffraction, it is the very element that makes protein crystal structures valuable. The high solvent content of protein crystals is essential for maintaining the structures of the macromolecules in their solution states. Therefore, to a large extent, proteins in crystals possess the structural, enzymatic, and functional properties of their counterparts in solution. Protein crystal structures must be regarded with care, however. In the hands of the uninformed, the danger exists that crystallographic structural data will be misinterpreted, or overreaching conclusions drawn. An understanding of the parameters derived from crystallographic experiments is essential if structural information from crystallographic experiments is to be used effectively to support drug discovery.

X-ray crystallography and light microscopy share the same basic principle; electromagnetic radiation scattered by the object to be imaged is recombined and focused by a lens to reform the image of the object. Theoretically, the resolving power of any imaging technique is equal to one half of the wavelength of the radiation used for imaging. To resolve the atomic details of protein structures, crystallographic experiments involve the exposure of protein crystals to high-energy monochromatic X-rays (wavelengths on the order of 1 Å). Imaging using X-rays is complicated by the fact

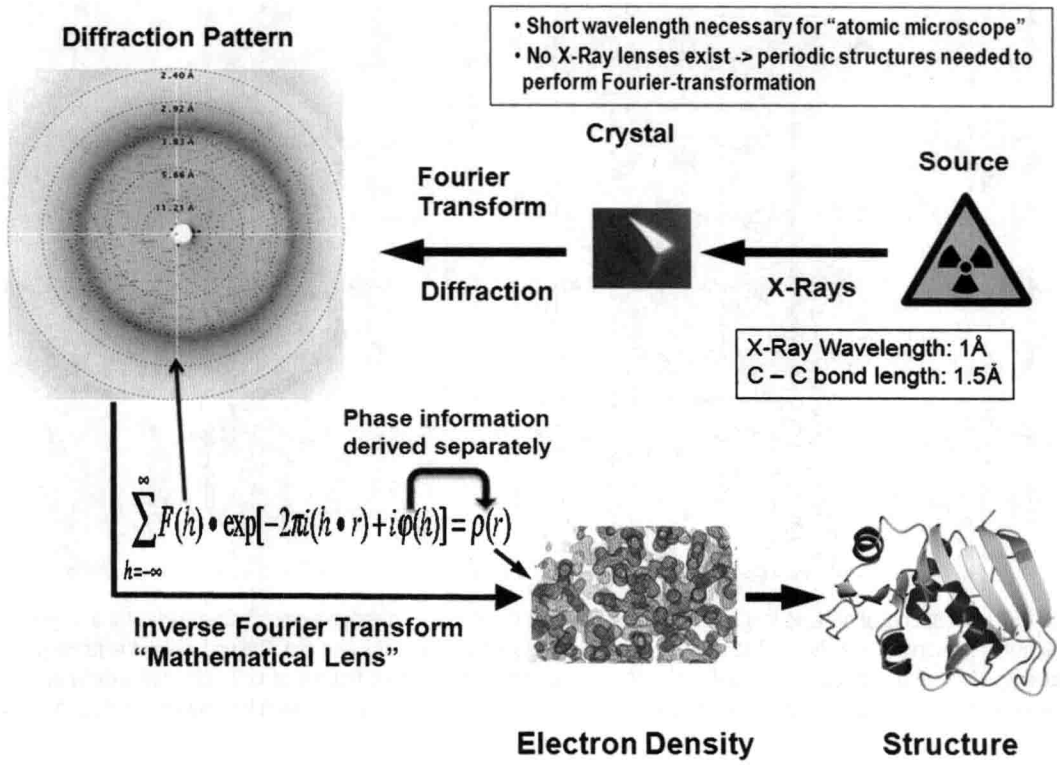


Fig. 2. A schematic outlining the steps in a crystallographic structure determination. Crystals are systematically exposed to monochromatic X-rays in multiple orientations, and the diffraction patterns are captured with electronic detectors. Since crystals are three-dimensionally periodic substances, the diffraction pattern comprises a series of spots rather than a continuous function. Each spot represents a family of diffracted waves that map to discrete spatial periodicities in the unit cell of the crystal. The diffraction pattern is a summation of waves of electromagnetic radiation and can thus be described by a Fourier series, and the diffraction pattern and disposition of the atomic contents of the unit cell are related mathematically by a Fourier transform. An image of the atomic contents of the unit cell of the crystal is derived by applying a mathematical lens (inverse Fourier transform, equation shown on the *lower left*) to the diffracted X-rays. The image reconstruction process is complicated by the fact that only intensities of the diffracted X-rays are measurable ($F(h)$ terms in the equation shown), but not the relative phase shifts between each family of diffracted waves. The missing information is referred to as the crystallographic phase problem. The missing phases are obtained using other experimental or computational methods described in the text. Since the diffraction of X-rays is caused by the interaction of the X-rays with electrons, the resulting image obtained in a crystallographic experiment is of the electron density distribution in the unit cell of the crystal. Interactive model building software is used to build the final atomic model into electron density.

that X-rays interact very weakly with matter, so that no lenses exist which are able to reconstruct the image from the scattered X-rays. Hence, the scattered X-rays from crystals must be captured with electronic detectors and the function of a lens must be simulated mathematically. A schematic describing the steps involved in the solution of a crystal structure is shown in Fig. 2.

Mathematical reconstruction of the structure of the atomic contents of the crystal is complicated by the fact that one of the two key pieces of information describing the diffracted X-ray waves, the relative phase shifts between the different families of diffracted

waves, cannot directly be measured (see Fig. 2). Three methods are commonly employed to overcome the phase problem, as summarized below.

(a) *Molecular replacement.* When an approximate model of the unknown crystal structure is available, it can be used to overcome the phase problem. The principle is simple; the model is first oriented and then positioned in the unit cell of the target crystal structure using rotation and translation functions. The correctly oriented model is subsequently used to calculate approximate phases and electron density maps. Alternate cycles of interactive correction and rebuilding of the model into electron density and model refinement are used to improve the quality of the phases and to transform the model structure into the real structure. The success of molecular replacement depends critically on two factors: the fraction of the asymmetric unit for which suitable models exist, and the r.m.s. deviation (after optimal superposition) between the model and target structures. Generally, r.m.s. deviation increases with decreasing sequence identity, or in cases where the target structure undergoes significant conformational changes with respect to the model structure (e.g., movement of protein domains). In the latter case, the model structure can be separated into individual fragments that are sequentially oriented and positioned in the unit cell. Newer maximum-likelihood molecular replacement algorithms, such as those implemented in the program Phaser (4) are more discriminating, and have been successful in solving difficult molecular replacement problems that were previously intractable.

(b) *Isomorphous replacement methods.* This is a classical approach used to solve protein structures with unknown folds. Crystals are soaked in multiple solutions containing salts of heavy atoms such as Hg, Pt, Pb, Au, etc., until conditions are found where a small number of heavy atoms incorporate in well-defined positions on the crystallized protein molecule (without altering the structure of the underlying protein). By analyzing the differences in the intensities of diffraction patterns from the native and heavy atom derivatized protein crystals, it is possible to determine the locations of the heavy atoms in the unit cell and to use the scattering "signal" from the heavy atoms to calculate phases and an electron density map (reviewed in refs. (5–7)).

(c) *Anomalous scattering methods.* For heavier elements, some inner shell electrons have absorption edges in the range of the X-ray wavelengths used in diffraction experiments. The heavy atoms in the protein crystal cause absorption of the impinging radiation, and impart small phase shifts on the radiation scattered from the crystal. This phenomenon is used to determine

the positions of the heavy atoms in the unit cell, and subsequently to extract phase information to allow electron density map generation. Anomalous scattering can be used to supplement the phase information obtained from isomorphous heavy atom derivatives, or to independently obtain complete phase information. A very powerful de novo phase determination method utilizes anomalous scattering from proteins that are homogeneously labeled with selenomethionine (incorporated during recombinant expression of the protein in *Escherichia coli*), a derivatized selenium-containing amino acid. Independent diffraction experiments are carried out (on the same crystal, if possible) at multiple X-ray wavelengths on the high and low energy sides of the selenium absorption edge that maximize the anomalous diffraction signal. This method requires a tunable X-ray source, which is present only at synchrotrons (reviewed in refs. (5–7)).

X-ray diffraction is caused by the interaction of the electric field vector of monochromatic X-rays with electrons in a protein crystal. These details, coupled with the fact that crystals are made up of three-dimensionally periodic lattices of molecules, have several important consequences (for excellent reviews see refs. (5–7)): (1) X-ray diffraction experiments generate three-dimensional images of the electron density distribution of the molecular components of the crystal. So heavier atoms generate a proportionally stronger signal, and hydrogen atoms are generally not discernable in protein crystal structures. (2) The short wavelength radiation used in X-ray diffraction experiments allows for the resolution of macromolecular structures at an exquisite level of detail (typical protein crystal structures are determined at resolutions between 1.5 and 3.0 Å resolution). (3) In a crystallographic experiment, the structure of the molecular contents of the unique portion of a crystal (called the asymmetric unit of the unit cell, which is the microscopic building block of the crystal) are obtained, and the resulting crystal can be built by the application of crystallographic symmetry operators to the contents of the asymmetric unit, as shown in Fig. 1. Since the diffraction signal from a crystal arises from constructive interference from trillions of crystallographic asymmetric units, the resulting crystal structure comprises a time- and space-averaged picture of the contents of the copies of asymmetric units that are sampled. Hence, components of the asymmetric unit with a large degree of random spatial heterogeneity, i.e., disordered protein loops or side chains and the bulk solvent occupying the spaces between protein molecules, fade into the background and cannot be modeled. However, in cases where a molecular component of a crystal, such as a protein side chain, occupies a finite number of distinct, low energy conformations in different asymmetric units, it is possible to simultaneously characterize each alternative conformation.

Examination of the equation relating diffracted X-rays to the crystal structure provides insight into the structural parameters that are modeled in a crystallographic experiment (see Eq. 1).

$$F_{hkl} = \sum_{j=1}^{N} f_j e^{-(B\sin^2\theta)/\lambda^2} e^{2\pi i(hx+ky+lz)}. \quad (1)$$

Equation 1 is one of the explicit forms of the structure factor equation (8). Each F_{hkl} term represents a unique family of diffracted X-ray waves from the crystal (diffracted waves from crystals constructively interfere to form patterns of spots, as shown in Fig. 2, which can each be assigned integer indices h, k, and l), which correspond to discrete spatial periodicities in the crystal lattice. The intensity and phase of each family of diffracted waves is derived via a summation of the scattering contributions from all of the atoms in the asymmetric unit of the crystal. The second exponential term in Eq. 1 computes the net phase shift relative to an arbitrary origin of the scattered wave with index h, k, l due to the relative positions of the individual atoms in the unit cell (with fractional coordinates x, y and z). The f_j term corresponds to the scattering factor for each atom in the summation, and is directly proportional to the number of electrons in the atom in question. The first exponential $B\sin^2\theta/\lambda^2$ term (θ is the angle of the scattered radiation with respect to the source X-ray beam, and λ is the wavelength of the X-rays) accounts for the reduction in the intensity of the scattered radiation with scattering angle due to interference between scattered waves from different parts of the electron cloud surrounding each atom. X-ray scattering is attenuated further by smearing of the electron clouds surrounding each atom due to thermal motion of the atoms. Atomic thermal motion is modeled using the extra B term in the structure factor equation. As a first approximation it is assumed that the thermal motion of atoms is isotropic (spherically symmetric), with $B=8\pi^2\mu^2$, where μ is the root mean square amplitude of atomic vibration. Using the calculation above, for a B-factor of 15 Å^2, the displacement of an atom from its equilibrium position is approximately 0.44 Å, and it is as much as 0.87 Å for a B-factor of 60 Å^2. Thus, analysis of B-factors is very important during any structural analysis to provide insight into the dynamics and structural integrity of different regions of a protein molecule. However, one must exercise caution before interpreting B-factors too quantitatively. In addition to measuring dynamic disorder caused by temperature dependent vibration of atoms, the B-factor is also influenced by subtle structural differences between protein molecules in different unit cells throughout the crystal (which spatially smears the atom positions), steric constraints from intermolecular lattice contacts, and certain systematic experimental errors, such as absorption of the X-ray beam during X-ray data collection. Advanced mathematical models can be used to provide more

detailed information on atomic thermal motions. For example, the relative motions of entire protein domains can be characterized using TLS refinement (9). Also, when high-quality X-ray data are available from crystals that diffract to high resolution (typically better than 1.2 Å, rare in protein structure determinations), the isotropic thermal correction can be replaced by a tensor, which corrects not only for the extent of thermal motion of the atoms but also for spatial anisotropy in their motions (10).

Based on the mathematical description of X-ray diffraction provided above, four parameters are optimized in a single crystal X-ray diffraction experiment for each atom in a protein crystal structure: the *x*, *y*, and *z* coordinates of each atom and the *B*-factor describing the thermal motion of each atom. The quality of resulting electron density maps and the accuracy of refined parameters in protein crystal structures are largely dependent on the resolution of the X-ray diffraction data (equivalent to the pixel size of electron density sections). Examples of the effects of diffraction resolution on electron density map quality are shown in Fig. 3. The model is generally manually built (or refit) into electron density by a crystallographer, using two types of electron density maps, $|2F_o - F_c|\alpha_c$ maps, and $|F_o - F_c|\alpha_c$ difference maps, described below.

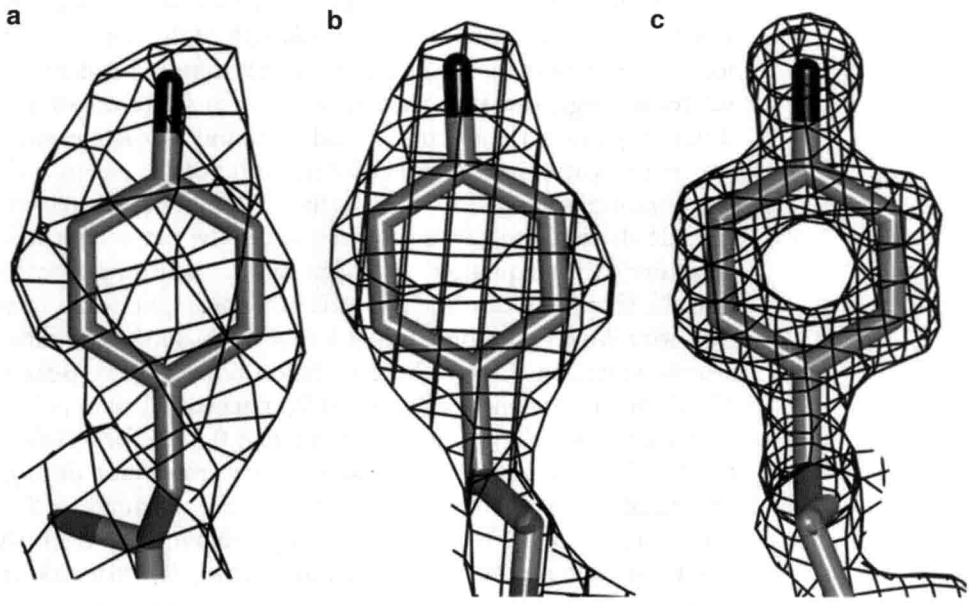


Fig. 3. Representative electron density maps contoured around tyrosine residues (using $|2F_o - F_c|\alpha_c$ coefficients) from three refined crystal structures: (**a**) A 2.8 Å resolution structure of *Francisella tularensis* topoisomerase IV, (**b**) A 2.2 Å structure of *Escherichia coli* topoisomerase IV, and (**c**) A 1.4 Å structure of *Enterococcus faecalis* DNA gyrase B (all from D. Bensen and L. Tari, unpublished results). The electron density maps were contoured using the electron density visualization software COOT (*see* ref. (11), Chapter 6). At better than 3.0 Å resolution, amino-acid side chains can be recognized with the help of protein sequence information, while at better than 2.5 Å resolution solvent molecules can be observed and added to the structural model with some confidence. As the resolution improves to better than 2.0 Å resolution, fitting of individual atoms may be possible and most of the amino-acid side chains can be readily assigned even in the absence of sequence information.

此为试读,需要完整PDF请访问: www.ertongbook.com