Max-Planck-Institut für Biochemie, Abteilung für Strukturforschung

Development of Theoretical and Experimental Methods of Time-Resolved Laue Protein Crystallography

Gleb Bourenkov

Vollständiger Abdruck der von der Fakultät für Chemie
der Technischen Universität München zur Erlangung des akademischen Grades

eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzender: Univ.-Prof. Dr. H. Schmidbaur

Prüfer der Dissertation:
1. apl. Prof. Dr. Dr. h.c. R. Huber
2. Univ.-Prof. Dr. F. G. Parak
3. apl. Prof. Dr. W. Bode , Ludwig-Maximilians-Universität München

Die Dissertation wurde am 19.09.2002 bei der Technischen Universität München eingereicht und durch die Fakultät für Chemie
am 05.11.2002 angenommen.

Abstract

New theoretical, experimental and computational methods for time-resolved Laue crystallography of macromolecules have been developed. A solution to the energy-overlap problem in Laue diffraction is described that does not require redundancy in the measurements. The new method follows a Bayesian approach with multi-dimensional probability density functions. The intensity components of reflection multiplets are deconvoluted, and estimates of their precision are obtained. The Laue patterns are processed to their physically relevant wavelength-dependent resolution limit; no "soft parameters" are involved. The power of the method is demonstrated by the test applications to bovine trypsin, sperm whale myoglobin and 2*Mn*-catalase from *Thermus thermophilus*. In the example of trypsin, the completeness at low and medium resolution as well as at very high resolution (1.4 Å) is enhanced very substantially as compared to standard procedures; the "low-resolution hole" problem is practically solved. As a consequence, the contrast in electron density maps improves so far that they become comparable in quality to maps from monochromatic data of the same resolution. The new method is of interest for all types of Laue diffraction experiments, in particular for single-shot time resolved studies on short time scales. In a stroboscopic test study of ground-state CO myoglobin, the Bayesian method permitted to obtain the electron density maps that showed the correct ligand conformation at resolution 1.7 Å. The resolution that was achieved using DORIS and the map contrast are higher as compared to the earlier experiments of this type that has been carried out at the third generation synchrotron source. The possibility to obtain interpretable maps in a single-shot Laue experiment is demonstrated for the example of cubic catalase. A limited-bandwidth Laue method using a graphite monochromator with 2.5% bandwidth has been developed and tested in study of lysozyme at ultra-high resolution (< 1.2 Å). A new algorithm for autoindexing Laue diffraction patterns has been developed and implemented in a computer program. Successful tests of this procedure on structures with relatively large cell constants >200 Å are presented.

Zusammenfassung

Neue theoretische, experimentelle und rechnerische Methoden für die zeitaufgelöste Laue-Kristallographie von Makromolekülen wurden entwickelt. Eine Lösung für das Energie-Überlappungsproblem der Laue-Diffraktion wird beschrieben, die keine Redundanz in den Messdaten erfordert. Die neue Methode nutzt Bayes-Statistik mit mehrdimensionalen Wahrscheinlichkeitsdichtefunktionen. Die Intensitätskomponenten der Reflexmultiplets werden entfaltet und ihre Genauigkeit abgeschätzt. Die Laue-Beugungsbilder werden bis zur jeweiligen physikalisch relevanten wellenlängenabhängigen Auflösungsgrenze bearbeitet, ohne dabei "Soft parameter" anzunehmen. Die Leistungfähigkeit der Methode wird anhand von Testanwendungen an Rindertrypsin, Pottwal-Myoglobin und 2*Mn*-Katalase des *Thermus thermophilus* demonstriert. Im Fall von Trypsin liegt die Vollständigkeit bei niedrigen und mittleren sowie bei sehr hohen Auflösungen (1.4 Å) im Vergleich zu Laue-Standardmethoden um ein Vielfaches höher. Das Problem der geringen Vollständigkeit im Bereich niederer Auflösung ("low-resolution-hole") ist praktisch gelöst. Als Folge erhöht sich der Kontrast der Elektronendichtekarten so weit, dass sie in ihrer Qualität mit Karten vergleichbar sind, die auf monochromatischen Daten gleicher Auflösung beruhen. Die neue Methode ist für alle Arten von Laue-Beugungsexperimenten, insbesonders auch für zeitaufgelöste Studien aufgrund einer einzigen Aufnahme (single shot) in kurzem Zeitmaßstab von Interesse. In einer stroboskopischen Teststudie an CO-Myoglobin im Grundzustand ermöglichte es die Bayes-Methode, Elektronendichtekarten zu erhalten, die die korrekte Konformation der Liganden bei einer Auflösung von 1.7 Å zeigten. Die Auflösung und der Kontrast der Dichtekarten, die aufgrund von Messungen an DORIS erhalten wurden, sind höher als bei früheren Experimenten dieser Art, obwohl diese an Strahlungsquellen der dritten Generation durchgeführt wurden. Die Möglichkeit, interpretierbare Karten mit Single-Shot-Laue-Experimenten zu erhalten, wird am Beispiel von kubischer Katalase demonstriert. Zudem wurde eine Laue-Methode mit reduzierter Bandbreite (2.5%) unter Verwendung eines Graphit-Monochromators entwickelt und am Beispiel von Lysozym bei extrem hoher Auflösung (< 1.2 Å) getestet. Schließlich wurde ein neuer Algorithmus zur Autoindizierung von Laue-Beugungsmustern entwickelt und in einem Computerprogramm implementiert. Erfolgreiche Anwendungen dieses Verfahrens auf Strukturen mit vergleichsweise großen Zellkonstanten werden vorgestellt.

Contents

LIST OF FIGURES V

LIST OF TABLES VI

INTRODUCTION 1

CHAPTER 1

THE WHITE BEAM LAUE METHOD 5

1.1 Diffraction geometry 6
Ewald construction 6
The spot shape 7

1.2 Integrated intensity 8

1.3 Scope of applications of the Laue method 10
1.3.1 Superior quality of monochromatic data 10
1.3.2 Time-resolved studies 11

CHAPTER 2

BAYESIAN ANALYSIS OF LAUE DIFFRACTION DATA 12

2.1 The "low resolution hole" and its effect on the contrast in electron density maps 12

2.2 Outline of the Bayesian method 15
2.2.1 Bayesian analysis of monochromatic X-ray diffraction data 15
2.2.2 The Bayesian approach for the case of Laue diffraction 16
2.2.3 An example illustrating Bayesian deconvolution 20
2.2.4 Relationship between least-squares and Bayesian solutions 23

2.3 Parameters influencing the accuracy in Laue data processing 24
2.3.1 Wavelength dependence in the apparent diffraction limit 24
2.3.2 The importance of accurate wavelength normalization 26

2.4 Implementation in a computer program 29

2.5 Further enhancement of the Bayesian method 29

CHAPTER 3

TEST APPLICATION OF THE BAYESIAN METHOD TO LAUE DIFFRACTION FROM TRYPSIN AT 1.4 Å RESOLUTION 33

3.1 BPT Laue data collection **34**

3.2 Detailed procedure of the data processing **34**

3.3 Results **36**

3.5 Contrast in Laue density maps in case of an incomplete model **41**

CHAPTER 4

APPLICATIONS OF THE BAYESIAN METHOD IN TIME-RESOLVED EXPERIMENTS **43**

4.1 Stroboscopic time-resolved studies of CO rebinding in myoglobin **43**
4.1.1 Data Collection and processing 45
4.1.2 Omit-maps of the CO ligand 47
4.1.3 Detection of reaction intermediates 49

4.2 Single-shot Laue experiments: Application to cubic *2Mn*-catalase **50**

CHAPTER 5

LIMITED-BANDWIDTH LAUE METHOD **54**

5.1 Experimental **56**
5.1.1 Monochromator setup 56
5.1.2 HEWL high-resolution data collection and processing 56

CHAPTER 6

AUTOMATIC INDEXING OF LAUE PATTERNS **63**

6.1 Description of the method **66**

6.2 Computer program LAIP **74**

6.3 Test applications **76**

DISCUSSION **78**

CONCLUSIONS **83**

REFERENCES **84**

LIST OF PUBLICATIONS **96**

List of figures

Fig. 1.1 Ewald construction for polychromatic Laue experiment **7**
Fig. 2.1 Comparison of the effects of sharpening and low-order truncation on the density distribution for the example of a one-dimensional structure model **13**
Fig. 2.2 Joint and marginal PDFsfor the intensity components of a double Laue reflection obtained by a single observation **18**
Fig. 2.3 Completeness and R_{mono} factors of BPT Laue data sets processed using wavelength-independent diffraction limits **25**
Fig. 2.4 Wavelength normalisation curve **28**
Fig. 3.1 Completeness of Laue data versus resolution **37**
Fig. 3.2 Standard and Bayesian Laue (2Fo-Fc) electron density maps for BPT- Tyr 59, Lys 60, Ser 61 and Gly 62 **39**
Fig. 3.3 Standard and Bayesian Laue (2Fo-Fc) electron density maps for BPT- Pro 124, Thr 125 and Ser 27 **39**
Fig. 3.4 Correlation factor from a comparison of Laue (F_o,φ_{calc}) density maps to a monochromatic map **40**
Fig. 3.5 Omit map of the BPT activation domain **42**
Fig. 4.1 Laue diffraction pattern of MbCO crystal obtained with 4000 individually extracted bunches **46**
Fig. 4.2 *Fo-Fc* electron density map in the active site of the MbCO **48**
Fig. 4.3 Laue diffraction pattern of cubic *2Mn*-Catalase **52**
Fig. 4.4 SIngle-shot Laue Electron density maps of TTC at 2.5 Å resolution **53**
Fig. 5.2 2.5%-bandwidth Laue diffraction patterns of lysozyme **58**
Fig. 5.3 The angular interval in the reciprocal space covered in a single Laue exposure with 2.5% bandwidth **60**
Fig. 5.4 2Fo-Fc electron density maps around Tyr55 residue **62**
Fig. 5.5 2Fo-Fc and Fo-Fc electron density maps of Cys76-Cys94 disulfide bond **62**
Fig. 6.1 Laue patterns of BPT (low density form, top) and yeast hexokinase p152k **64**
Fig. 6.2 Geometries for deriving equations for autoindexing
Fig. 6.3 Laue pattern of yeast hexokinase PII after rapid diffusion of glucose **77**

List of Tables

Table 2.1. Example of the Bayesian deconvolution **22**
Table 3.1. Comparison of BPT Laue data to monochromatic data **38**
Table 4.1. MbCO Laue data processing statistic **47**
Table 5.1. HEWL limited-bandwidth Laue data statistics **61**
Table 6.1. Experimental conditions and results of autoindexing **75**

Introduction

Transient changes in the three-dimensional structure of proteins are essential for their biological function. Extending the methodology of protein crystallography and diffractometry towards structure determination of substates with short (seconds to pico-seconds) live times may significantly contribute to a better understanding of the mechanisms of protein function. Laue diffraction from protein crystals permits the simultaneous exploration of a large volume in reciprocal space. Under favourable conditions characterised by a broad incident x-ray wavelength distribution, a high-symmetry space group, and high crystalline order, the simultaneously excited reflection spots may yield structure factor amplitudes for a high percentage of all possible independent reflections. Due to the high spectral brightness in the hard X-ray wavelength range of a number of synchrotron radiation sources, including in particular wigglers with short critical wavelengths, such a Laue diffraction pattern may be recorded within very short exposure times. This provides a potential for time-resolved investigation of protein molecules in structural states with lifetimes in the μs - s regime. Even single-bunch exposures with an intrinsic time resolution <100 *p*s are feasible (Szebenyi et al., 1992, Wulf et al., 1997).

Revival of the old Laue method (Amoros et al., 1975) as a means of quantitative structure analysis with synchrotron radiation begun in the mid-eighties (Moffat, Szebenyi & Bilderback, 1984, Helliwell 1984, 1985). Theoretical analyses of Laue geometry were given by Cruickshank and coworkers (1987, 1991). Data processing software has been developed and applied to model systems by several groups

(Helliwell et al., 1987, Smith Temple, 1987, Bartunik & Borchert, 1989, Wakatsuki, 1993). A number of static Laue studies were reported, including applications to microcrystals (Hedman et al., 1985) and a virus (Campbell et al., 1990). Substrates, inhibitors and metal binding sites were identified in a number of feasibility studies (Hajdu et al., 1987a, Hajdu et al., 1987b, Farber et al., 1988, Reynolds et al., 1988, Campbell et al., 1990). Complete structure determinations by molecular replacement with Laue data were reported (Howell et al. 1992, Vellieux et al., 1993). First studies with time-resolution on the scale of seconds on ras-P21 (Schlichting et al., 1989), γ-chymotrypsin (Stoddard et al., 1990) and trypsin (Singer et al., 1993) were used by the authors to draw a function-related conclusions.

Together with potential promises, these pioneering studies revealed a number of fundamental and practical problems of the Laue technique. One fundamental problem was caused by a lack of low-resolution data due to energy overlapping of reflections. This "low-resolution hole" caused errors in Fourier calculations, hence low contrast and quality of electron density maps (Bartunik, Bartsch & Qichen Huang, 1992; Andersson et al., 1992). The other major drawback has been related to an extreme sensitivity of the method to crystalline disorder. Critical assessments of the technique led to the conclusion that *"as it stands, the Laue method is not the method of choice for time-resolved structural studies"* (Hajdu & Andersson, 1993). New approaches to the problem were necessary.

A number of subsequent developments solved the fundamental probem if the low-resolution hole. Ren & Moffat (1995a) and Bourenkov et al. (1996) developed methods that made it possible to deconvolute energy overlaps and to obtain electron density maps of high quality from Laue data. The least-squares method by Ren & Moffat (1995a) is based on the highly redundant measurement of multiple reflections. This method was applied in a study of carbon monoxide (CO) rebinding following photolysis of CO-myoglobin (MbCO), which demonstrated the feasibility of *n*s time-resolved protein crystallography. Later time-resolved applications exploiting the Laue method focused on studying photo-stimulated cyclic reactions on a nanosecond time scale using a "pump-probe" approach (Srajer et al., 1996, Perman et al., 1998). In such experiments it may be possible to collect a large number, e.g. 40, Laue exposures at different crystal orientations. Radiation damage on the one hand and

the need for many exposures at different crystal orientations on the other result in relatively poor resolution. The data were limited in resolution, e.g. to ~2 Å in experiment on the photolysis of MbCO (Srajer et al., 1996).

The results of ultra-high resolution (<1.2Å) studies of static structures of the initial and final states of the ligandation reaction (CO- and deoxy-myoglobin, Kachalova et al., 1999) revealed that the binding of the CO molecule to the heme iron is in fact accompanied by conformational changes. Such changes, involving atomic displacement < 0.4 Å, however, could not be detected in these previous time-resolved studies because of inadequate resolution.

The scope of this thesis is the development of new approaches aiming to improve the data quality and to reach higher resolution. The development and the experimental verification of a fundamentally new method of deconvoluting energy (and spatial) overlaps in Laue diffraction patterns from protein structures forms the central part of the thesis (Chapters 2,3). The method follows a Bayesian probabilistic approach using Wilson (1949) statistics as a source of *a-priory* available information. The use of this method permits to fill the "low-resolution hole". Test applications showed dramatic improvements in the contrast and the quality of electron density maps. Furthermore, the method made it possible to extend the data from trypsin to 1.4 Å resolution - the by far highest resolution ever reached in Laue diffraction studies of proteins. The use of the Bayesian method in the processing of nanosecond time-resolved Laue diffraction data from CO-myoglobin produced structural information that extended to 1.7 Å resolution despite a small number of crystal settings and the relatively low flux density on BW6 at DORIS (Chapter 4). Furthermore, the method does not require redundancy in the measurement of multiple reflections; therefore, for the first time single-shot Laue experiments became feasible. The first successfull application of this possibility is demonstrated in Chapter 4 for the example of cubic 2Mn-catalase from *Thermus thermofilus*. The problems that are related to broadened crystal mosaicity may be solved by a "*limited-bandwidth Laue method*", which is described in Chapter 5. This method provides an alternative to the white-beam Laue method in pump-probe diffraction experiments. The method circumvents most of the problems of Laue technique (spatial and energy overlaps, high background, radiation damage), but requires a higher number of crystal orientations. Experiments were

performed using a double-crystal graphite monochromator with 2.5% bandwidth; the limited bandwidth method can also be applied to time-resolved studies at ultra-high (~1 Å) resolution using undulator radiation. The last Chapter, 6, is devoted to the development of a method of autoindexing Laue patterns from structures with large cell dimensions.

Chapter 1
The white beam Laue method

This chapter combines a description of fundamental aspects of the Laue method with a critical discussion of the parameters that are relevant for deriving accurate structure factors.

We define the *Laue diffraction pattern* as the diffraction pattern observed from a stationary single crystal exposed to a quasi-parallel X-ray beam with a wavelength bandwidth exceeding the intrinsic bandwidth of the Bragg reflections:

$\Delta\lambda_b > \Delta\lambda_c$. (1.1)

The subscripts *b* and *c* refer to "beam" and "crystal" correspondingly. $\Delta\lambda_c$ is defined by the mosaicity μ of the crystal. For the case of small mosaicity, $\mu << \Theta$ (Bragg angle) Relationship between $\Delta\lambda_c$ and μ is obtained by differentiation of the Bragg formula:

$\Delta\lambda_c/\lambda = \mu /\tan \Theta$. (1.2)

The mosaicity of protein crystals that are suitable for structure analysis lies in the range from 0.01° to 1°.

$\Delta\lambda_b$ is defined by the properties of the synchrotron source, the insertion devices and the beamline optical elements (mirrors, mono- or polychromators, windows, air path and detector).

Two types of the method should be distinguished, namely the *white beam Laue method*, when $\Delta\lambda_b >> \Delta\lambda_c$, and the *limited bandwidth Laue method*, $\Delta\lambda_b \sim \Delta\lambda_c$. In this chapter, the basic principles of the white beam Laue method are summarized.

1.1 Diffraction geometry

Ewald construction

Fig. 1.1 shows the Ewald construction for the case of the white-beam Laue geometry. Two Ewald spheres with radii $1/\lambda_{min}$ and $1/\lambda_{max}$ are indicated. All Bragg reflections corresponding to diffraction vectors (or reciprocal lattice points, RLPs, represented by the dots) lying between the two spheres are excited simultaneously. Diffracted rays corresponding to RLPs lying between the line OX and the$1/\lambda_{min}$ sphere will be recorded on a detector with an acceptance angle $2\Theta_{max}$. Assuming a round flat area detector positioned normally to the incident beam (downstream from the sample), the number of excited reflections within the geometrical resolution range ∞ to $d_{min}=\lambda_{min}/2\sin\Theta_{max}$ is

$$N= \frac{\pi}{6V^*}\left(\frac{1}{\lambda_{min}^3}-\frac{1}{\lambda_{max}^3}\right)\sin(3\Theta_{max}) \quad . \qquad (1.3)$$

V* is the reciprocal unit cell volume of the crystal.

If a radial line OY passes through a number, n(hkl), of RLPs, all the corresponding reflections project onto exactly the same spot on the detector. Such reflections are called harmonic or energy overlaps. We will refer to such a superposition of reflections as a multiplet of the order *n*. Only RLPs which lie between the circles $1/\lambda_{min}$ and $1/2\lambda_{min}$ may be recorded as singles; all RLPs between $1/n\lambda_{min}$ and $1/(n+1)\lambda_{min}$ are *n*-fold multiplets. The number of spots in a Laue pattern is 7/8 of the number of recorded reflections. Clearly, most of the low-resolution reflections will be buried in multiplets. The problem of resolving multiplet intensities into component reflections, i.e. harmonic deconvolution, is discussed in Chapter 2.

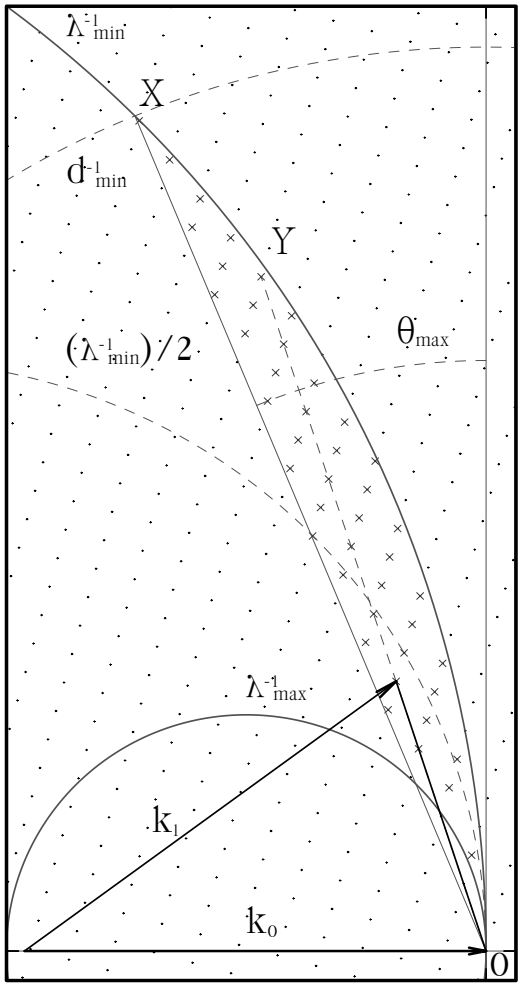

Fig. 1.1. Ewald construction for polychromatic Laue experiment.

The spot shape

The divergence of the beam diffracted by a crystal with mosaic spread μ is $m_{\parallel}=2\mu+\eta$ in the plane of diffraction, and is $m_{\perp}=(\mu+\eta)\sin(\Theta)$ in the perpendicular plane. Here η is the primary beam divergence in the corresponding direction; $\eta < 0.1°$ for focused synchrotron beam. For a given RLP, the radial and tangential sizes of the resolution function on a flat detector positioned at a distance D are:

$X_r=2D\ m\ \cos(2\Theta)^{-2}$

X_t=D m sin(Θ) cos(2Θ)$^{-1}$.

The reflection spot size is the result of a convolution of the resolution function with the projection of the sample (~100 μm) and the detector point spread function (~100 μm). The latter two parameters practically define the spot size and shape for the case when μ ≤ 0.1°. When μ>0.3°, the contribution of the mosaic spread dominates; and the spots have the shape of radial streaks with increasing length towards higher diffraction angles. Estimates of the integrated intensities from spatially overlapping spots may be obtained through modelling of reflection profiles (Shrive et al., 1992, Rhen & Moffat, 1995a).

1.2 Integrated intensity

The intensity which is integrated over a (single) Laue reflection *hkl* with interplanar spacing *d* and measured at wavelength λ, is related to the modulus of the structure factor *F* as follows (Zachariasen, 1945):

$$I_{hkl} = I_o(\lambda)\left(2d\lambda\frac{e^2}{mc^2Vo}\right)^2 PV\,|F_{hkl}|^2. \qquad (1.4)$$

Here P is a polarization factor (Kahn et al., 1982), V_o and V are the volumes of the unit cell and the crystal, respectively. $I_o(\lambda)$ is defined by a number of wavelength-dependent instrumental factors, including the source spectrum, detector efficiency, absorption etc. In order to estimate the $|F_{hkl}|$ values from measured integrated intensities, the wavelength normalization function $g(\lambda)=\lambda^2 I_o(\lambda)$ should be determined. The normalization procedure is discussed in Chapter 2.2.

The formula (1.4) does not take into account a variation of g(λ) over the reflection bandwidth. I discuss in the following the limits of this standard approach. Let us assume a crystal that consists of mosaic blocks, with μ^2 being the dispersion of the distribution function of the block (mis-)orientations. The scattering by each of the blocks obey equation (1.4). Assuming μ<<Θ, the crystal reflectivity as a function of the wavelength is described by a normalized function q(λ-λ') with the mean value λ and

dispersion $\Delta\lambda_c$ (equation 1.1). The integrated intensity (omitting the constants) is then given by:

$$I_{hkl} \propto Pd \mid F_{hkl} \mid^2 \int_{\lambda_{\min}}^{\lambda_{\max}} g(\lambda')q(\lambda - \lambda')d\lambda', \qquad (1.5)$$

In a Laue experiment, the exact form of $q(\lambda)$ is unknown. Therefore, (1.4) is used for estimating the structure factor moduli in practice. One can estimate the magnitude of the relative error introduced by this approximation as a first term in power series expansion of the difference between the $|F_{hkl}|$ value that is estimated according to (1.4) and the correct value:

$$\frac{\delta(\mid F_{hkl} \mid)}{\mid F_{hkl} \mid} \approx \frac{1}{2g(\lambda)^2} \int_{\lambda_{\min}}^{\lambda_{\max}} \left[g(\lambda) + (\lambda - \lambda')g'(\lambda) + \frac{1}{2}(\lambda'-\lambda)^2 \frac{\partial^2 g}{\partial \lambda^2} \right] q(\lambda'-\lambda)d\lambda' \qquad (1.6)$$
$$= \mu^2 (d^2 - \frac{\lambda^2}{4}) \frac{g''(\lambda)}{g(\lambda)}$$

The error may be quite big for low resolution reflections in the case of a crystal with high mosaicity, and/or when the source spectrum does not represent a smooth function (e.g. U46/ESRF undulator, Bourgeois et al., 2000). A theoretically calculated $g(\lambda)$ distribution for the wiggler beamline BW6/DORIS is shown in Fig. 2.4. Excluding the ranges proximal to the Au absorption edges and those close to λ_{min} and λ_{max}, the value |g"/g| is of the order of 70 Å^{-2}. When we request the error to be smaller then10% and redefine the mosaicity (in degrees) as a full width of a reflection[1], $\mu' \cong 3.3\ \mu$, then

d_{max} [Å] < 1/μ' [°]

gives an estimate of the low resolution limit at which the equation (1.4) is still applicable. Consequently, in order to measure low-resolution reflections with $d > d_{max}$, the function $q(\lambda)$ has to be modelled reliably.

[1] In the rotation method, the angular range of intensity integration, that provides the best signal-to-noice ratio (Lehman & Larsen, 1974), is ±1.65μ.

1.3 Scope of applications of the Laue method

Extended discussions of the general limitations of the white beam Laue method (in particular due to the mosaicity, cell volume vs. resolution, energy overlaps, “low resolution hole”, polychromatic background and long-wavelengths componets of the spectrum) may be found in the literature (e.g. in “’Time-resolved Macromolecular Crystallography”, edited by Cruickshank, Helliwell and Johnson (1992).).

1.3.1 Superior quality of monochromatic data

For static experiment using a given crystal and detector, monochromatic methods will always be superior to Laue diffraction, unless fast energy-resolving area detectors will become available. At present, claims that the Laue data were "at least as good" or "even better" (Yang et al., 1998, Ren et al., 1999) imply that monochromatic data selected for comparison were not of optimal quality.

A number of authors proposed to use Laue method for static structure determination in order to circumvent the problem of radiation damage at room temperature (Helliwell, 1991, Vellieux, 1993, Ren et al., 1999). The idea is based on the assumption that radiation damage is not only dependent on the X-ray dose absorbed by the sample, but also on the time which is elapsed since receiving the dose (Nave, 1995). However, in contrast to this assumption, the kinetic behaviour of the radiation damage as a function of the X-ray dose was found to be very similar when observed on a time scale of microseconds (Moffat et al., 1986) as compared to days (Sygush & Allaire, 1988). At photon flux levels bellow 10^{12} photon s^{-1} mm^{-2}, data are available (Popov, 2000) which demostrate the dependence of the radiation

damage on the X-ray dose only. At higher fluxes, the dependence may be different due to heating of the crystal (Glaeser et al., 2000). Because of the polychromatic background, the total dose needed to measure reflection intensities with Laue method to the same accuracy may be by up to several orders of magnitude higher then under monochromatic conditions.

1.3.2 Time-resolved studies

The advantages of the Laue method over the classical oscillation method in crystallographic data collection are as follows: i) a large number of reflections is measured simultaneously and ii) the reflection intensities are integrated over the energy instantaneously without changing the crystal orientation. Thus, the Laue technique is of potential interest to time-resolved studies, i.e. for measuring the diffraction data from crystals in substates with life-times that are shorter then the time required for monochromatic data collection. These include (i) studies of (irreversible) enzymatic reactions at room temperatures on time scales ranging (currently[2]) from microseconds to a few minutes; or (ii) studies of reversible reactions on time scales of the synchrotron bunch length (>100 *p*s). A number of theoretical and experimental method, which are of relevance to such applications, have been developed within the framework of the present work and are discussed in the following chapters.

[2] In a nearest future sub-seconds time-resolved monochromatic data collection will become feasible with pixel-array detectors (Xoung et al., 2000).

Chapter 2
Bayesian Analysis of Laue Diffraction Data

2.1 The "low resolution hole" and its effect on the contrast in electron density maps

The problem of harmonic overlaps is of fundamental nature. A considerable percentage (>20%) of all measurable reflection spots in Laue diffraction patterns are multiplets containing components from several different wavelengths. If the multiplets are not adequately deconvoluted into component intensities, the completeness in the diffraction data is substantially reduced, in particular at low to medium resolution. This "low-resolution hole" causes errors in Fourier calculations, hence low contrast and quality of electron density maps (Bartunik, Bartsch & Qichen Huang, 1992; Andersson et al., 1992). A detailed discussion of such effects in terms of real-space point spread functions is given by Duke et al. (1992), or in Patterson space by Samudzi and Rosenberg (1992).

The effects of the truncation of low reflection orders on the density distributions differ from density sharpening operations, that involve a deliberate down-weighting of low-resolution reflections. This is illustrated in Fig. 2.1 for the example of

a one-dimensional structural model containing widely separated density peaks. Unlike sharpening, the low-frequency perturbation due to a truncation of reflection orders at low resolution clearly reduces the contrast in the density distribution.

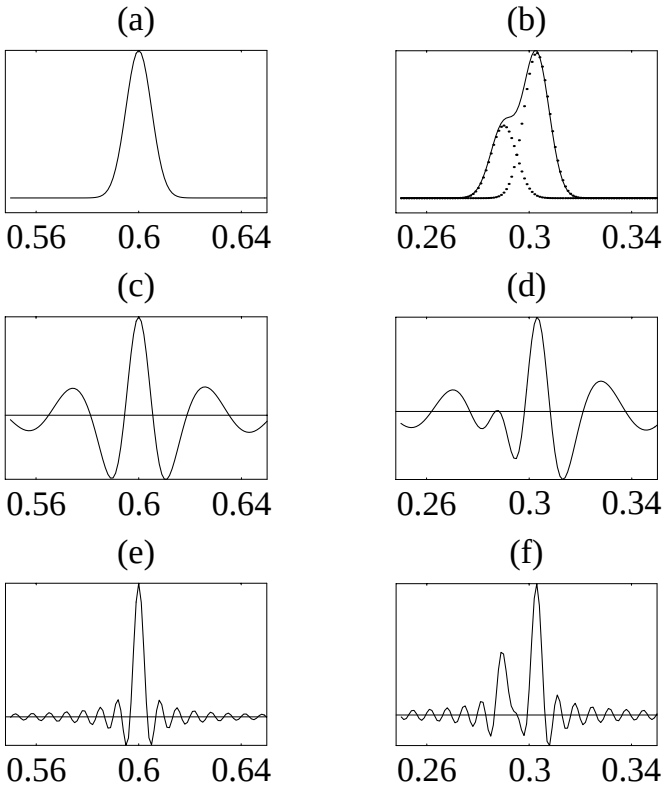

Fig. 2.1. Comparison of the effects of sharpening and low-order truncation on the density distribution for the example of a one-dimensional structure model. (a),(b) Original periodic density segments containing a single Gaussian peak (a) or a peak consisting of two overlapping Gaussians (b). The variance of the Gaussians is u=1/200 of the period. The diffraction pattern is calculated for the reflection orders h=0 to h=150. (b),(c) Reconstructed density after omitting reflection orders h=1 to h=30. (c),(d) Sharpened density, obtained by weighting the structure factors with the function $Exp[2(\pi hu)^2]$.

Maps that are affected by the low-resolution hole may be improved by including a small number of model structure factor amplitudes (Bartunik, Bartsch & Qichen Huang, 1992). However, as these authors discussed, such a procedure would introduce a bias in the interpretation of the maps, which is not acceptable considering the aim of time-resolved Laue diffraction studies to detect unknown conformational changes. In the specific case of a structure with high non-crystallographic symmetry, Hol and coworkers (Vellieux et al., 1993) were able to obtain good electron density maps through six-fold density averaging, despite low completeness (37%) in their Laue data. In general, however, without appropriate deconvolution of overlaps,

Fourier maps calculated with Laue structure factors necessarily must be of poor quality.

A number of authors previously developed techniques aiming to reduce or solve the energy overlap problem. Helliwell and co-workers (1989) suggested a method based on the use of a stack of photographic films and the dependence of the absorption in each film on the X-ray wavelength; the power of the method was limited by the poor energy resolution of photographic film (Wakatsuki, 1993). Another approach based on direct methods was proposed by Hao et al. (1993); the authors demonstrated the applicability to small-molecule structures. A closely related method involving a Patterson function modification (squaring) was suggested for application to protein structures (Hao, Harding & Campbell, 1995a). This method produced roughly correct (i.e., better than random) estimates of multiplet component intensities; no measure of the uncertainty of the estimates could be obtained. Substantially more accurate estimates may be obtained from least-squares deconvolution of harmonics based on the measurement of symmetry-related reflection multiplets (Helliwell, 1992; Wakatsuki, 1993; Campbell and Hao, 1993; Bradbrook et al., 1995, Ren and Moffat, 1995b). However, these methods require substantial redundancy in the observations. Furthermore, they have to be located in suitable ranges of the wavelength normalization curve. For example, if multiplet observations belong to an approximately linear region in the distribution of the normalization factor vs. wavelength, the resulting least-squares systems tend to degenerate. The need for high redundancy in practice makes it necessary to collect Laue data at additional crystal settings. This excludes single-shot experiments and represents a severe drawback in all applications involving short time scales.

All previously proposed methods of deconvoluting harmonics have in common that no solid justification is provided for treating a given reflection spot as single or multiple. In each of these procedures, the decision is made assuming a wavelength-independent diffraction limit as a "soft parameter" that is empirically chosen to best predict spots that are visible in the Laue pattern (e.g., Hao, Harding & Campbell, 1995b). This is adequate in the case of monochromatic diffraction patterns. In Laue diffraction, however, this way of defining a diffraction limit has no physical justification. In fact, the incident X-ray intensity, the diffracting power of the crystal and other

experimental factors vary with the wavelength. Therefore, the average signal-to-noise for each resolution shell and thus the apparent diffraction limit clearly must be wavelength dependent. This was first demonstrated when analyzing Laue exposures from yeast hexokinase PII (Bartsch et al., 1990; Rupp, 1991). In this study, the apparent diffraction limit was determined by comparing observed to predicted patterns and found to vary between 2.5 Å (for λ = 0.95 Å) and 3.2 Å (λ = 0.65 Å); the wavelength dependence was in good agreement with a theoretical estimate taking the signal-to-noise into account. Neglecting the wavelength dependence and assuming a universal soft diffraction limit leads to systematic under- or overprediction of the Laue pattern. Wrong assignment of the components of a given reflection, for example omission of some higher harmonics, introduces a bias; it affects both the completeness and the accuracy of data in all resolution shells.

We developed a fundamentally new method of deconvoluting energy (and spatial) overlaps in Laue patterns from protein structures. The method provides a complete and justified probabilistic solution. For each unique reflection that may be excited under the given experimental conditions defined by the wavelength bandpass and the geometries of the diffraction set-up and the detector, the best estimates of the structure factor modulus, its square, and their standard uncertainties are obtained. The results are independent from the multiplicity of the reflection; redundancy in the observation is not required. No soft parameters are involved. The only assumption made is the validity of Wilson statistics (Wilson, 1949). The new method follows a Bayesian statistical approach.

2.2 Outline of the Bayesian method

2.2.1 Bayesian analysis of monochromatic X-ray diffraction data

The use of probability theory following a Bayesian approach in the analysis of raw X-ray diffraction data was first proposed by Kheiker & Nekrasov (1970) for the case of monochromatic single-crystal diffraction. A similar concept formed the basis of the program TRUNCATE (French & Wilson, 1978) in the CCP4 program suite. These

developments aimed to find the best estimates of the structure factor amplitude F and it's square I from the intensity observation J, when J is small as compared to the noise or even negative. A-priori given information about I being positive and obeying Wilson's distributions was taken into account through probability density functions (PDFs),

$$p(X) \propto \exp(-X/\langle I \rangle) \qquad (2.1a)$$

$$p(X) \propto X^{-1/2} \exp(-X/2\langle I \rangle). \qquad (2.1b)$$

(2.1a) corresponds to acentric reflections (having phase φ, $0 \leq \varphi \leq 2\pi$), (2.1b) to centric reflections (having restricted phase $\varphi = n\pi/2$, n=0,1,2,3). The symbol $\propto$ denotes 'follows'. X is an I-associated random variable, and <I> is a mean intensity at given resolution.

Experimental observation of the reflection J provides parameters (J,σ_J) of the normal probability density $p(J|X)$ for the intensity. This information must be associated with the prior PDF p(X) trough the Bayes theorem, which states:

$$p(X \mid J) \propto p(X)\, p(J \mid X), \qquad (2.2)$$

and can be read as follows: *the posterior PDF, i.e. the probability of the observation J due to elementary hypothesis X follows the probability of the hypothesis multiplied with the probability of the hypothesis to produce the observation*. The best estimates for F, I, and their uncertainties are given by the moments of the resulting posterior PDF.

2.2.2 The Bayesian approach for the case of Laue diffraction

The previously developed Bayesian method (Kheiker & Nekrasov, 1970) is limited to non-overlapping reflections and independent observations. In Laue diffraction, an identical approach may be applied to non-overlapping single reflections

only. Analysis of Laue reflection multiplets, however, requires a Bayesian approach that deals with *multi-dimensional PDFs*.

An *n*-multiple Laue reflection intensity is given by the experimental observation in the form of

$$J = \mathbf{aX}. \tag{2.3}$$

Here $\mathbf{a}=\{a_i\}$ is a vector whose elements contain products of the Lorentz, polarization and wavelength normalization factors for the *n* X_i components of the variable vector **X**. If several exposures are made, the products further contain overall scale and temperature factors for frame-to-frame scaling. The components of **a** are always positive. We assume that the uncertainties of a_i are negligible as compared to the observation uncertainty σ_J; this is further discussed below.

Assuming a normal distribution of J, the observation of a reflection multiplet yields a degenerate multivariate normal PDF over the *n*-dimensional space of **X**:

$$p(J \mid \mathbf{X}) \propto \exp(-(\mathbf{aX}-J)^2/2\sigma^2). \tag{2.4}$$

Fig. 2.2(a) shows an example of such a function in the case of two variables. If a set of $m \geq 1$ redundant observations $\mathbf{J}=\{J_1,...,J_m\}$ is present, insertion of their product in the Bayes formula will produce a more generalized form of the normal PDF (2.4):

$$p(\mathbf{J} \mid \mathbf{X}) = \prod_{i=1}^{m} p(J_i \mid \mathbf{X}) \propto \exp[-\frac{1}{2}(\mathbf{X}-\mathbf{R})^T \mathbf{C}(\mathbf{X}-\mathbf{R})] \tag{2.5}$$

The expression contains the normal $n*n$ matrix $\mathbf{C}=\mathbf{A}^T\mathbf{WA}$. The rows of the matrix **A** are formed by $\mathbf{a}_j$ (j=1, ..., *m*). **W** is a diagonal matrix with elements $W_{jj} = \sigma_{Jj}^{-2}$. The vector **R** may be found as any vector which satisfies the equation

$$\mathbf{CR} = \mathbf{A}^T\mathbf{WJ}. \tag{2.6}$$

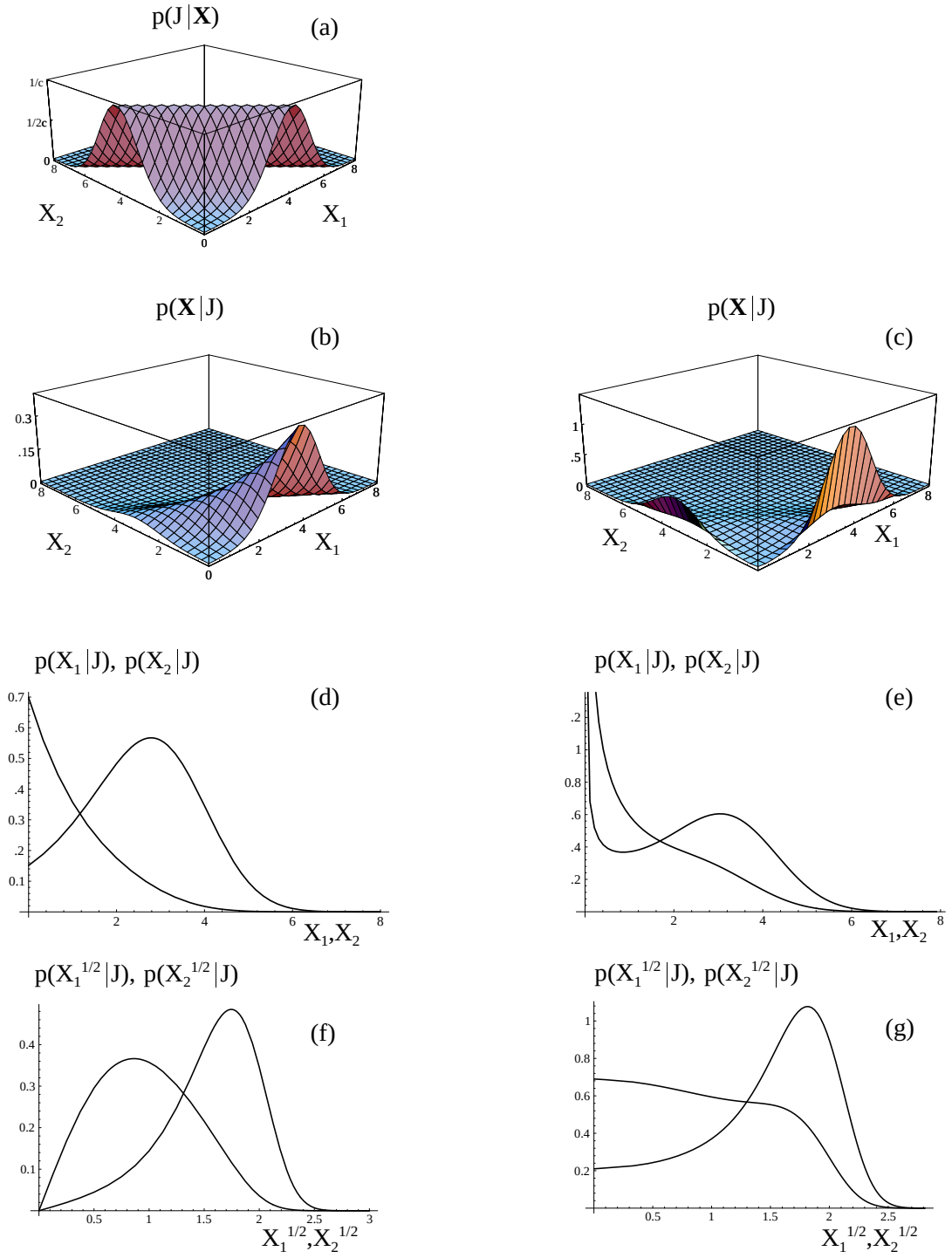

Fig. 2.2. (a) Joint PDF P(J|X) for the intensity components of Laue reflection with multiplicity 2 obtained by a single hypothetical observation with intensity J=5, σ_J=1. (b),(c) - Joint posterior PDFs with acentric (b) or centric (c) Wilson priors $\langle I \rangle_1$=3, $\langle I \rangle_2$=1. (d),(e) - Marginal posterior PDFs for intensities in the acentric (d) or centric (e) case. (g),(f) - Marginal posterior PDFs for amplitudesin the acentric (g) or centric (f) case.

Due to the positivity of the components of **X**, the resultant PDF is a truncated multivariate normal over the positive hyperquadrant of the **X**-space. This PDF is never degenerate, even if the **C** matrix is. In a next step, we associate this PDF through the Bayes theorem with an appropriate Wilson prior for the components and obtain one of the following posterior PDFs on the positive hyperquadrant:

$$p(\mathbf{X} \mid \mathbf{J}) \propto \exp[-\mathbf{hX}-\frac{1}{2}(\mathbf{X}-\mathbf{R})^{T}\mathbf{C}(\mathbf{X}-\mathbf{R})], \tag{2.7}$$

for a set of acentric reflections, $h_i=1/\langle I\rangle_i$, and,

$$p(\mathbf{X} \mid \mathbf{J}) \propto \prod_{i=1}^{m} X_i \exp[-\mathbf{hX}-\frac{1}{2}(\mathbf{X}-\mathbf{R})^{T}\mathbf{C}(\mathbf{X}-\mathbf{R})], \tag{2.8}$$

for set of centric reflections, $h_i=1/2\langle I\rangle_i$. Figs. 2.2 (b), (c) show p(**X**|**J**) for the example of two variables. Within a given Laue multiplet, the reflection components are either all centric or all acentric.

When the *n*-dimensional PDF is found, the resulting posterior probability density for the component X_k is the projection of p(**X**|**J**) onto the (p,0,X_k) plane:

$$p(X_k \mid \mathbf{J}) \propto \int_0^{\infty}\cdots\int_0^{\infty} p(\mathbf{X} \mid \mathbf{J})\, dX_1 \cdots dX_{k-1}\, dX_{k+1} \cdots dX_n \tag{2.9}$$

(Figs. 2.2 (d), (e)). The best estimates for the structure factor amplitude $(F_o)_k$, it's square $(I_o)_k$, and their standard uncertainties are given by the moments of the posterior PDF (2.9):

$$F_{0k} = \int_0^{\infty} X_k^{1/2}\, p(X_k \mid \mathbf{J})\, dX_k \tag{2.10}$$

$$I_{0k} = \int_0^{\infty} X_k\, p(X_k \mid J)\, dX_k \tag{2.11}$$

$$\sigma^2_{F_{0k}} = I_{0k} - F^2_{0k} \tag{2.12}$$

$$\sigma^{2}{}_{I_{0k}} = \int_{0}^{\infty} X_k^2 \, p(X_k \mid \mathbf{J})\, dX_k - I_{0k}^2 \qquad (2.13)$$

2.2.3 An example illustrating Bayesian deconvolution

Let us consider the example of an acentric reflection with multiplicity 2 and only one observation. This situation refers to a great part of the Laue data in an experiment, which is aimed to obtain maximum crystallographic information from a minimum number of crystal settings. The PDF (2.9) in this case has the form (Fig. 2.2 (d)):

$$p(X_1) \propto \exp\left[\left(\frac{1}{<I>_1 a_1} - \frac{1}{<I>_2 a_2}\right)\frac{a_1 X_1}{\sigma_J}\right]\left(1 + \mathrm{Erf}\left[\frac{1}{\sqrt{2}\,\sigma_J}\left(J - \frac{\sigma_J^2}{<I>_2 a_2} - X_1 a_1\right)\right]\right) \qquad (2.14)$$

The estimates (2.11,2.13) yield exact analytical expressions:

$$(I_o)_1 = \frac{\sigma_J}{a_1}\left(\Delta_{12} + \frac{1+\xi_1\chi_1}{\chi_1 - \chi_2}\right), \qquad (2.15)$$

$$\sigma^2_{(I_o)_1} = \frac{\sigma_J^2}{a_1^2}\left[2\Delta_{12}^2 + \frac{2(1+\xi_1\chi_1)\Delta_{12} + (1+\xi_1^2)\chi_1 - \xi_1\chi_1}{\chi_1 - \chi_2}\right] - I_{o_1}{}^2, \qquad (2.16)$$

were $\xi_i = J/\sigma_J - \sigma_J/<I>_i a_i$, $\chi_i = (\pi/2)^{1/2} \exp(\xi_i^2/2)\{1 + Erf(\xi_i/2^{1/2})\}$, $\Delta_{ij} = (1/<I>_i a_i - 1/<I>_j a_j)^{-1}$.

The correlation coefficient is:

$$c_{12} = \left(\frac{\sigma_J^2}{a_1^2 a_2^2}\left[2\Delta_{12}^2 + \frac{(2+\xi_1\chi_1+\xi_2\chi_2)\Delta_{12}}{\chi_1 - \chi_2}\right] - I_{o_1} I_{o_2}\right)\left(\sigma_{I_{o_1}}\sigma_{I_{o_2}}\right)^{-1/2}. \qquad (2.17)$$

In the special case $<I>_1 a_1 = <I>_2 a_2$ ($\xi_1 = \xi_2 = \xi$, $\chi_1 = \chi_2 = \chi$) the formulae (2.15)-(2.17) transform in the following way:

$$(I_o)_i = \frac{\sigma_J}{a_1}\frac{(1+\xi^2)\chi+\xi}{2(1+\xi\chi)}, \tag{2.18}$$

$$\sigma^2_{(I_o)_i} = \frac{\sigma_J^2}{a_1^2}\frac{2+\xi^2+(3\xi+\xi^2)\chi}{3(1+\xi\chi)} - I_{o_i}^{\ 2}, \tag{2.19}$$

$$c_{12} = \left(\frac{\sigma_J^2}{a_1 a_2}\frac{2+\xi^2+(3\xi+\xi^2)\chi}{3(1+\xi\chi)} - I_{o_1} I_{o_2}\right)\left(\sigma_{I_{o_1}}\sigma_{I_{o_2}}\right)^{-1/2}. \tag{2.20}$$

In order to illustrate the effect of the analysis, the posterior moments for a series of hypothetical observations are listed in Table 2.1. The values of J/σ_J range from -3 to 100, of $\langle I\rangle_2 a_2$ from 1 to 20; $\langle I\rangle_1 a_1$ was chosen to be 20.

When the observed sum of component intensities is small, one notices as a rather trivial result that both terms are small and the estimates are practically uncorrelated. The case of $J/\sigma_J \geq 3$ is more interesting. For the first term, which a-priori is supposed to be stronger, one gets useful estimates even when the ratio $\langle I\rangle_2 a_2 / \langle I\rangle_1 a_1$ is as high as ~75%. But this does not mean at all that the second term may be ignored, even not when $\langle I\rangle_2 a_2 / \langle I\rangle_1 a_1$ is only ~ 5% , since this would cause a shift in the estimates by more then a standard deviation.

In most cases, the procedure affects both terms of a doublet. The posterior moments for the weaker term in general provide less accurate information, depending on the relative size of the components and on the signal-to-noise. Only in the case when $\langle I\rangle_2 a_2 \ll \langle I\rangle_1 a_1$ and $J \gg \sigma_J$, no information at all is obtained about the second term (i.e., the posterior moments do not significantly differ from the prior ones). Of course, there is always a finite probability that the true value of the intensity of the second term is even higher then the intensity of the first one. This is expressed by a joint posterior PDF, that results in a modified posterior expected value and an increase in the standard uncertainty.

When $\langle I\rangle_2 a_2$ approaches the value $\langle I\rangle_1 a_1$, i.e. when prior preference may not been given to any of the components, then evidently the probability for both components to have any intensity between 0 and J/a_i will be uniform. As $J/\sigma_J \to \infty$, $I_{oi} \to J/2a_i$ and $I_{oi}/\sigma_{Ioi} \to 3^{1/2}$. For the components of the real Laue doublet, a substantial

Table 2.1. Posterior expected values, standard uncertainties, and correlation coefficients for the intensities of components derived with Wilson's prior from a set of hypothetical observations of a double acentric Laue reflection with unit standard uncertainty. The prior for the first component is assumed $\langle I \rangle_1 a_1 = 20$. In each cell, the moments are arranged in the scheme:

$a_1(I_o)_1 \quad a_1\sigma_{a1(Io)1}$

$a_2(I_o)_2 \quad a_2\sigma_{a2(Io)2}$

c_{12}

Observation	Prior for the second component $\langle I \rangle_2 a_2$									
J	1		5		10		15		20	
-3	0.27	0.25	0.26	0.25	0.26	0.25	0.26	0.25	0.26	0.25
	0.21	0.21	0.25	0.24	0.26	0.25	0.26	0.25	0.25	0.24
	-0.048		-0.055		-0.056		-0.057		-0.057	
0	0.66	0.54	0.62	0.52	0.62	0.52	0.62	0.52	0.62	0.52
	0.43	0.39	0.58	0.50	0.60	0.50	0.61	0.51	0.62	0.62
	-0.170		-0.204		-0.209		-0.211		-0.212	
3	2.21	1.1	1.75	1.1	1.68	1.1	1.66	1.1	1.64	1.1
	0.84	0.73	1.47	1.0	1.59	1.1	1.62	1.1	1.64	1.1
	-0.521		-0.618		-0.626		-0.628		-0.630	
6	4.93	1.38	3.49	1.85	3.21	1.86	3.11	1.86	3.06	1.86
	1.02	1.0	2.56	1.77	2.89	1.83	3.00	1.85	3.06	1.86
	-0.694		-0.852		-0.857		-0.859		-0.859	
10	8.90	1.45	6.21	2.83	5.44	2.95	5.16	2.96	5.03	2.96
	1.05	1.05	3.78	2.75	4.59	2.92	4.88	2.95	5.03	2.96
	-0.732		-0.937		-0.943		-0.943		-0.943	
20	18.9	1.45	14.3	4.81	11.6	5.67	10.6	5.79	10.0	5.8
	1.05	1.05	5.61	4.72	8.35	5.64	9.44	5.78	10.0	5.8
	-0.725		-0.937		-0.984		-0.985		-0.985	
50	48.9	1.45	43.3	6.64	34.4	12.5	28.4	14.2	25.0	14.4
	1.05	1.05	6.54	6.56	15.5	12.5	21.6	14.2	25.0	14.4
	-0.725		-0.989		-0.997		-0.998		-0.998	
70	68.9	1.45	63.3	6.73	52.1	15.6	41.6	19.5	35.0	20.2
	1.05	1.05	6.66	6.66	17.8	15.6	28.3	19.5	35.0	20.0
	-0.725		-0.989		-0.998		-0.999		-0.999	
100	98.9	1.45	93.3	6.74	80.6	18.2	63.3	27.0	50.0	28.9
	1.05	1.05	6.67	6.67	19.3	18.2	36.7	27.0	50.0	28.9
	-0.725		-0.989		-0.998		-0.999		-0.999	

difference between priors may always be expected due to the nature of the wavelength normalization curve and the dependence of $\langle I \rangle$ on $\sin\theta/\lambda$. This forms the main difference of our approach to the Bayesian approach which Sivia & David (1994) proposed for an analysis of overlapping reflections in monochromatic powder diffraction patterns. In their method, there is no difference in the priors; thus, use of the uniform positive prior is permitted.

2.2.4 Relationship between least-squares and Bayesian solutions

When redundant observations are available (e.g., at different crystal settings) and **C** is a well defined matrix, (6) has a unique solution **R**. The posterior PDF for acentric reflections (2.7) and the exponential term in (2.8) both convert to a multivariate normal through a shift of the centre to the point

$$\mathbf{R}' = \mathbf{R} - \mathbf{C}^{-1}\mathbf{h}. \qquad (2.21)$$

If all values $R'_i/(\mathbf{C}^{-1})_{ii}^{1/2} > 3$, then (2.10) - (2.13) can well be expressed by an expansion of the integral to the whole **X**-space, and $(F_o)_k = R'^{1/2}_k$, $(I_o)_k = R'_k$, $\sigma_{(Io)k} = (\mathbf{C}^{-1})_{kk}^{1/2}$, $\sigma_{(Fo)k} = (\mathbf{C}^{-1})_{kk}^{1/2} / 2(F_o)_k$ represent fairly good approximations to the moments. In this case our approach converges to the least squares solution, which takes the components of **R** as the best estimates of intensities with the covariance matrix $\mathbf{C}^{-1}$ (if proper weights are used).

The use of the least squares method becomes problematic not only when **C** is degenerate, but also when some of the R_i (at least one) are negative. An application of the TRUNCATE procedure (French & Wilson, 1978) to the estimates that are extracted in such a way would be completely invalid. Wakatsuki (1993) and Campbell & Hao (1993) proposed to solve the problem by eliminating outlying terms. Nonlinear solution of (2.6) for the $R_i^{1/2}$, as proposed by Ren & Moffat (1995b), yields identical results. From the point of view of the underlying PDFs, these methods use the coordinates for a point where the truncated PDF (2.5) adopts its maximum as estimates for the intensity, and their square roots as estimates for the amplitude. Setting a negative observation to zero introduces a maximal bias to the data (French & Wilson, 1978), even in the case of independent measurements. In the case of a multiple reflection, negativity of R_i for a component does not necessarily imply that the corresponding term is negligible. Rather, it may indicate the existence of strong

correlation between the variables; under such conditions, the least-squares system is poorly defined. The joint PDF in this case is similar as in the case of a completely degenerate system, which is combined with an acentric prior (Fig. 2.2(b)). In both cases, although the weaker term has a maximum-likelihood value at zero, the contribution of the corresponding component to the observed intensity is significant. Eliminating this component will yield systematically overestimated values and underestimated standard uncertainties of the intensities of the other terms.

In the powder diffraction method of Sivia & David (1994) which only considers the amplitudes, the estimates were proposed to be taken as the maxima of the PDF (2.4) after transformation to the amplitude space. We stress the point that in general the best estimates are not provided by the maxima of the PDF, but by its moments. For example, if we consider a Fourier transform of the resulting data, then its value at each point yields an expected value (approaching the maximum likelihood value through the central limiting theorem) of the electron density only in the case when the first moments are taken as coefficients.

2.3 Parameters influencing the accuracy in Laue data processing

2.3.1 Wavelength dependence in the apparent diffraction limit

Protein Laue data evaluation by standard procedures involves an uncertainty in the choice of the diffraction limit. Before the development of the Bayesan approach, it has never been reported that a protein Laue data set was processed to the actual geometrical diffraction limit $d_{min}=\lambda_{min}/(2\sin\theta_{max})$. Rather, a "soft" diffraction limit at poorer resolution was chosen. An analogous procedure was proposed by Hao, Harding and Campbell (1995b) for introducing a "soft" minimum wavelength λ_{min}, which may exceed the experimentally defined value. The assumption of a constant

diffraction limit causes under- or overprediction of the Laue patterns; thus, it may strongly affect the results of the evaluation (Bartunik, Bartsch and Huang Quichen, 1992).

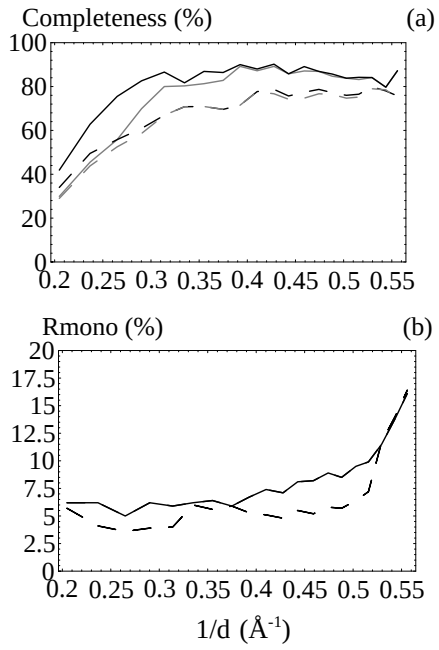

Fig. 2.3. Completeness (a) and R_{mono} factors from comparison to monochromatic data (b) of BPT Laue data sets processed using wavelength-independent diffraction limit at 1.8 Å (solid lines) and at 1.3 Å (dashed lines). The completeness for single reflections is indicated by grey lines. R_{mono}, i.e. R-factor of scaling to monochromatic data, is shown for singles only.

Using our test data (see Chapter 3 for details), one can demonstrate the effect of a truncation that is introduced by choosing a wavelength-independent diffraction limit. Careful visual analysis of the patterns showed reflections up to about 1.8 Å. We processed the data with the Daresbury Laue Software Suit (Campbell, 1993) assuming a constant d_{min} value of 1.8 Å and 1.3 Å, respectively. The results are presented in Fig. 2.3. One can clearly see from the processing to 1.8 Å that omission of 'invisible' high resolution reflections yields data of lower accuracy. On the other hand, processing to 1.3 Å resolution with a standard (non-Bayesian) procedure results in a dramatically reduced completeness, since more reflections are designated as multiples which can not be deconvoluted.

The Bayesian approach solves the problem of wavelength dependence in the apparent diffraction limit in a most general way. The formulation involves the parameter $<I>a/\sigma_J$, which in fact is a wavelength dependent signal-to-noise ratio. If the pattern has been evaluated using an obviously overestimated diffraction limit (e.g., the geometrical diffraction limit) and unshifted estimates taken from (2.10)-(2.13), not a single reflection will be lost due to overprediction. Furthermore, no bias is introduced in each particular observation. In the case when the contribution of a high resolution component to a given multiplet intensity is negligible, then both Wilson's priors (2.1 a) and (2.1 b) approach the $\delta^+(0)$ function, and the posterior moments for the strongest term approach the expressions derived by Kheiker & Nekrasov (1970) for the case of a single reflection.

The estimates and their uncertainties will produce a modified signal-to-noise ratio for each reflection or for entire resolution shells. This ratio may be used in the same way as with monochromatic data for truncating data at an appropriate resolution or for excluding individual reflections.

2.3.2 The importance of accurate wavelength normalization

All formulae derived in Chapter 2.2.2 are based on the assumption that the coefficients a_i are known accurately. Explicit consideration of errors in these coefficients leads to a highly multi-dimensional (n^2m instead of n) integral in the calculation of the moments. In order to minimize the effects of these errors and to improve the accuracy for the whole data set, the wavelength normalization factors should be as precise as possible. When this is achieved, the remaining influence of the errors in the a_i may be neglected. This is demonstrated by the example given below.

Previously developed wavelength normalization techniques used a wavelength binning method, followed by power series smoothing (Helliwell et al.,

1989) or modelling by Tchebychev polynomials (Smith Temple, 1989, Arzt et al., 1999). However, if the wavelength normalization is based on measurements of symmetry equivalent single reflections only, high redundancy in the measurements (e.g., at different crystal settings) is needed for achieving sufficient accuracy. Furthermore, there are difficulties in extending such normalization curves to the long X-ray wavelength range. If a reflection from a plane with spacing d is observed at $\lambda > 2\lambda_{min}$, then the reflection spot on the detector contains a further contribution corresponding to the plane spacing d/2, unless there is systematic extinction. The contribution of the second harmonic might be negligible, and thus, the reflection spot effectively single. However, this may only be verified if both the ratio of the mean intensities at d and d/2 and the ratio between the wavelength normalisation factors at λ and $\lambda/2$ are known. Extrapolation of the polynomial to higher λ values is not justified, since its coefficients are physically meaningless.

An alternative approach is to derive the wavelength normalization curve from a reference crystal, not necessarily for the same structure, and to scale the Laue data to external monochromatic data. Multiple reference reflections may be included. We subdivide the wavelength range into a suitable number of bins. Reliable bin scale factors require a sufficient number (not less then say 40) reference data points in each bin. The bin scale factors g_i are obtained by least-squares solution of equations of the general form $J^{ref}=\Sigma I^{mono}{}_i g_i$. The reference patterns should be processed with an overestimated wavelength range; the effective limits λ_{min} and λ_{max} are defined on the basis of the resulting wavelength normalization curve. Fig. 2.4 shows an example of a normalization curve that was derived in this way from measurements on the beamline BW6 at DORIS. The discontinuities in the curve correspond to absorption edges of heavy atoms that are contained in the beam-defining components of the instrument (Au coated planar and toroidal mirrors; Fe contamination of Be windows) and the imaging plate used as a detector (Ba, Br, Eu). Wavelengths in the close vicinity of such discontinuities were omitted during the binning procedure. Except for these bins, the curve perfectly matches to the theoretical calculations. Corrections for wavelength

dependent absorption by the sample crystal and its environment relative to the

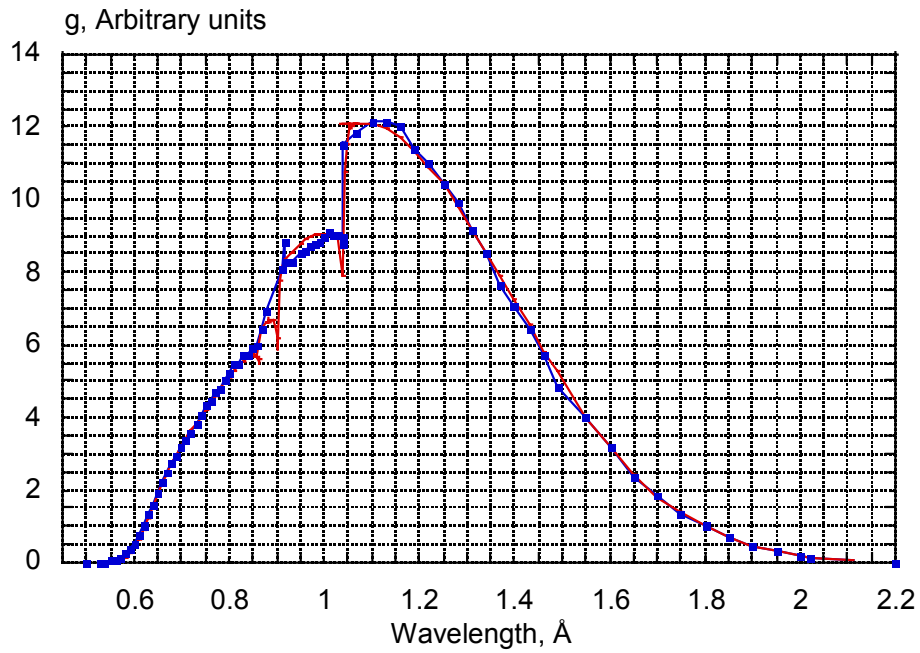

Fig. 2.4 Wavelength normalisation curve (blue) calculated from scaling to reference monochromatic data, and a theoretical curve (red) calculated from the BW6 wiggler spectrum (30 mm gap) and the transmission of two (planar and toroidal) gold-coated mirrors with 4 *m*rad glancing angle. In order to correct for absorption in the beam path, sample and capillary, the theoretical curve was scaled by a function $\mathrm{Exp}(-\alpha\ \lambda^3)$. The coefficient α was fitted by least-squares.

reference crystal are applied by fitting a single parameter. If the sample contains a resonant absorber with an absorption edge within the incident bandwidth, two (or more) parameters have to be fitted for two (or more) different segments of the entire wavelength range.

2.4 Implementation in a computer program

The formalism derived in Chapter 2.2.2 was implemented in a computer program. The estimates for the structure factors and the intensities are derived according to eqs. (2.10)-(2.13). Integration is carried out numerically, since we could not derive an analytical expressions analogous to (2.15-2.20) for the general case. This part of the procedure is the most complicated step. A special algorithm was developed for sampling the multi dimensional PDFs (2.7,2.8), which permits to integrate the moments in a reasonable time with a precision of at least two significant digits in up to ten dimensions. The verification of the method in an application to experimental data is described in the next chapter.

2.5 Further enhancement of the Bayesian method

The previously discussed method is based on the validity of Wilson distributions, i.e. on the assumption that atoms are uniformly distributed over the unit cell. The estimates of the structure factor moduli are independent of the structural model that provides the phases. The task may be reformulated in terms of finding the best estimates of the Fourier coefficients given a set of Laue multiplet observations and a structural model. For the case of single observations (i.e. monochromatic data), the problem was first considered by Srinivasan and co-workers (Srinivasan & Ramachandran, 1965, Srinivasan & Chandrasekharan, 1966), and further investigated by Bricogne (1982, 1993, 1997). It is closely related to the problem of maximum-likelihood structure refinement (Pannu & Read, 1996, Murshudov et al., 1997).

More accurate Fourier coefficients could be obtained with the following modified approach.

In time-resolved studies, the protein structure will be known approximately. For example, when investigating a reaction in the crystal, the structure corresponding

to the initial state will have been determined previously. In an acentric case, an *a-priory* knowledge of a structural model is expressed in terms of the conditional probability density on a complex plane C_y, were *y* is the random variable associated with the structure factor *F* of a reflection from the current structure:

$$P(y \mid F_c) \propto \exp(-\mid y - F_c \mid^2 / \varepsilon\sigma^2{}_c) .$$

Here F_c is a (complex) structure factor representing the model. Parameter σ_c has a sense of the average difference $|F_c\text{-}F|$ at a given resolution. In the case of monochromatic data, the value of σ_c is evaluated by maximum-likelihood methods (Read, 1986) based on a subset of reflections (e.g. those excluded from the model refinement (R-free subset) (Lunin & Skovoroda, 1995)). In a Laue case σ_c may be determined by the same procedures from a subset containing only single reflections. In general, σ_c is expected to be small as compared to the average $|F|$ value in time-resolved Laue experiment (Ursby & Burgeois, 1997), e.g. when a highly accurate model of the initial state is available and small conformational changes are being studied.

The multivariate Gaussian PDF (2.5) defined by the set of N-multiple Laue observations is converted into the y-space :

$$\mathrm{p}(\mathbf{J} \mid \mathbf{yy}^{cc}) \propto \exp[-\frac{1}{2}(\mathbf{yy}^{cc}\text{-}\mathbf{R})^{T}\mathbf{C}(\mathbf{yy}^{cc}\text{-}\mathbf{R})],$$

A posterior PDF for complex structure factors has the form:

$$\mathrm{p}(\mathbf{y} \mid \mathbf{J}) \propto \exp[-\frac{1}{2}(\mathbf{yy}^{cc}\text{-}\mathbf{R})^{T}\mathbf{C}(\mathbf{yy}^{cc}\text{-}\mathbf{R}) - \mathbf{s} \mid \mathbf{yy}^{cc} - F_c F_c^{cc} \mid^2] .$$

Here $s_i = 1/\varepsilon\sigma_i^2$. The best estimates of Fourier coefficient for a reflection is the first moment of the posterior PDF:

$$< F_k >= \int_{Cy_1}\!\!\int\!\!\ldots \int_{Cy_N}\!\!\int y_k \; \mathrm{p}(\mathbf{y} \mid \mathbf{J})\, dy_1 dy_1^{cc} \ldots dy_N dy_N^{cc} .$$

The integrals that are necessary in order to calculate the posterior moments may be expressed in terms of arguments z_i via the transformation to polar coordinates, $y_i=z_i(\cos\varphi_i+I\sin\varphi_i)$ and integration over the phases φ_i:

$$\left| \int_{Cy_1}\!\!\int\!\!\ldots \int_{Cy_N}\!\!\int \mathrm{p}(\mathbf{y} \mid \mathbf{J})\, dy_1 dy_1^{cc} \ldots dy_N dy_N^{cc} \right| =$$

$$\int_0^{\infty}\!\!\ldots\int_0^{\infty} \exp\left[-\frac{1}{2}(\mathbf{z}^2 - \mathbf{R})^{\mathrm{T}} \mathbf{C} (\mathbf{z}^2 - \mathbf{R}) - \mathbf{s}\mathbf{z}^2\right] \prod_i z_i \, \mathrm{Io}(2 s_i F_{c\,i} z_i)\, dz_1 \ldots dz_N$$

$$\left| \int_{Cy_1}\!\!\int\!\!\ldots \int_{Cy_N}\!\!\int y_k \; \mathrm{p}(\mathbf{y} \mid \mathbf{J})\, dy_1 dy_1^{cc} \ldots dy_N dy_N^{cc} \right| =$$

$$\int_0^{\infty}\!\!\ldots\int_0^{\infty} \exp\left[-\frac{1}{2}(\mathbf{z}^2 - \mathbf{R})^{\mathrm{T}} \mathbf{C} (\mathbf{z}^2 - \mathbf{R}) - \mathbf{s}\mathbf{z}^2\right] z_k^2 \, \mathrm{Io}(2 s_k F_{c\,k} z_k) \prod_{\mathrm{i}\neq\mathrm{k}} z_i \, \mathrm{Io}(2 s_i F_{c\,i} z_i)\, dz_1 \ldots dz_N$$

The Fourier coefficients obtained in this way, particularly the values $<F>_k - F_{ok}$, which are used for a difference synthesis, should be considerably more accurate as compared to those obtained with a Wilson's (1949) prior. However, in a situation when a multiplet observation contains no information on a particular component (as it was discussed in Chapter 2.3), the posterior moments will be substituted by the prior ones. Care should be taken in interpretation of such maps which may be biased towards the model.

Numerical calculation of the integrals above is considerably more complicated as compared to (2.10-2.13) and will require development of new algorithms. As a possible simplified solution one may use the maximum of the function under the first integral as an estimate of the first moment. In fact, this function represents a posterior distribution of the structure factor moduli, and has a unique non-trivial maximum on a positive hyperquadrant of the z-space. Similar approximation is being successfully used in powder diffraction methods (Sivia & David, 1994), even when uniform PDF is taken as a prior. As an alternative, one may consider approximation of the prior by a

Gaussian function expressed in the X-space. Such an approach has been proposed for modelling nonisomorphism for use in experimental phasing against intensities (Lebedev et al., 1998). In this case, the posterior PDF converts to a multivariate Gaussian form (2.6,2.7), and may be integrated using an available numerical algorithm.

Chapter 3
Test application of the Bayesian method to Laue diffraction from trypsin at 1.4 Å resolution

The Bayesian method was tested using Laue data from benzamidine inhibited ß-trypsin from bovine pancreas (BPT) as a model structure. The crystal form is orthorhombic (space group $P2_12_12_1$), with cell dimensions a=54.8, b=58.4, c=67.8 Å. The crystal structure was previously refined at 1.6 Å resolution (Marquart et al.,1983), and further refined at 1.0 Å resolution (Popov & Bartunik, 1996). The latter BPT model and monochromatic intensity data are used for comparison throughout the chapter.

The data processing procedure presented in this chapter outlines all steps of a complete protocol of Laue data evaluation making use of the Bayesian approach. The results will be compared to those obtained using a "standard" procedure, which is essentially corresponds to the data processing protocol suggested by the Daresbury Laue Software Suite (Campbell, 1993).

3.1 BPT Laue data collection

Laue data were collected on the double-focusing wiggler beamline BW6 at DORIS using a MAR imaging plate detector (diameter 300 mm). Ten exposures were recorded from a single crystal at a temperature of 10 ^{o}C. For the first exposure, the crystal was oriented with its a-axis close to the beam direction and the b-axis close to the spindle axis. Subsequent exposures were taken at 10^{o} intervals in the spindle rotation angle. A wavelength normalization curve was derived from a reference data set. The wavelength range was determined to $0.52 \leq \lambda \leq 2.1$ Å. The crystal mosaicity was about 0.1^{o}. The detected Laue reflection spots had a size of ≈400 μm (FWHM) in both the radial and tangential directions.

3.2 Detailed procedure of the data processing

The processing of the data started by refining the crystal cell, orientation and detector setting parameters with LAUEGEN (Campbell, 1993). A final prediction was made assuming a geometrical diffraction limit d_{min}=1.0 Å. Spot integration was performed with the program INTLAUE (Shrive et al.,1990). Negative observations were kept throughout the processing. Considering spatially overlapping reflections, we treated observations as uncorrelated if they resulted from a deconvolution of reflections that were separated by more then one FWHM. The intensities of the reflections with centroid-to-centroid distances less than one FWHM were added up and further treated in the same way as multiplets.

The exponent index α of the absorption correction coefficient of the form $\exp(\alpha\lambda^3)$, frame-to-frame scale factors and the temperature factors were fitted as individual parameters for each of the exposures on the basis of redundant measurements of single reflections. Scaled singlet data were divided into appropriate resolution ranges, and <I> values were calculated for the ranges d > 1.27 Å. Low completeness (47%) of the data in this step in general does not strongly affect the accuracy in the <I> values. A number of about 40 reflections in each bin is sufficient

to get reliable estimates (French & Wilson, 1978). This is in good agreement with the theory of structure-factor statistics (Stanley, 1955). At d < 1.27 Å, the <I> values were strongly affected by a poor signal-to-noise ratio. Therefore, we used an extrapolation assuming that <I> in the high resolution range followed the distribution in the mean squared scattering factor (Wilson, 1949). We interpolated the dependence over the range 2.0-1.27 Å and then extrapolated to 1.0 Å resolution. Prior moments for the intensities at d < 2.0 Å were obtained by linear interpolation between the centres of neighbouring bins.

In view of the fact that overprediction may introduce additional noise (not a systematical bias) in the case of non-deconvolutable spatial overlaps, predicted reflections were removed if the prior probability of the reflection intensity being higher than the noise did not exceeded 1%. An apparent resolution limit was obtained that varied, for $\lambda=\lambda_{min}$, between 1.10 Å for the first to 1.22 Å for the last pattern; this reflects an increase in the overall temperature factor by 1.3 Å² during the entire data collection. Spot reintegration was then repeated with a correspondingly modified prediction of the pattern.

In the further processing, all observations were treated in the same way, whether they were single or multiple. The redundant observations were found and the data were divided into subsets of cross-correlated reflections in such a way that each unique reflection appeared in one subset only. Normal matrices **C** were calculated for each subset, vectors **R** were found through the singular values decomposition of **C**, and **h** were formed with the prior.

In the program TRUNCATE (French & Wilson, 1978), single reflections are considered as negligible if $[J-(\sigma_J{}^2 h)] \leq -4\,\sigma_J$. By analogy, the intensity contribution of the k-th component of a multiple reflection may be treated as negligible when $[(\mathbf{A}^T\mathbf{W}\mathbf{J})_k - h_k/C_{kk}] \leq -4\,C_{kk}^{-1/2}$. Thus, we omit components on the basis of an estimate of an “upper limit" of its significance. After that, the data are input into the deconvolution program which calculates (numerically) the expected values and standard uncertainties for the intensities and structure factor moduli of all remaining unique reflections, that appear in the multiplets of the order ≤ 10.

3.3 Results

Bayesian estimates of the intensities and the structure factors were obtained for subsets containing up to ten unique reflections; they included multiplets up to order 10. Resolution shells were rejected if $\langle I \rangle / \langle \sigma_{Io} \rangle \leq 2$. This limited the resolution of the final data set ("Bayesian Laue data", BLD) to 1.4 Å. Further reflections were omitted if the estimates of σ_{Io} exceeded $\langle I \rangle$; this limit corresponds to the unmodified prior. BLD contained in total 36,044 reflections.

In order to illustrate the power of the new method, we produced an alternative data set ("standard Laue data", SLD) following a conventional procedure. A constant "soft" resolution limit of 1.65 Å was used; use of a higher resolution limit would have led to even more serious overprediction problems. The same scale and wavelength normalization factors were taken as for BLD. All multiple reflections were included that could be processed with a least squares procedure. This yielded a total number of 20,796 reflections, corresponding to 78% of the recorded unique reflections. At the same nominal resolution of 1.65 Å, BLD included 95% of the recorded reflections.

The completeness of both data sets is compared in Fig. 3.1. The completeness improves substantially over the whole resolution range. At low and medium resolution ($d > 3.3$ Å) it increases from 55% to 84%, and at high resolution ($d > 1.65$ Å) from 77% to 90% in going from the standard processing to the Bayesian procedure. The completeness drops to or below 50% (Bradbrook et al., 1995) at $d \geq 6.3$ Å and 3.7 Å for BLD and SLD respectively. In addition, the Bayesian approach provides a substantial number of reflections at high resolution $1.65 \text{ Å} \geq d \geq 1.4$ Å (completeness 68%).

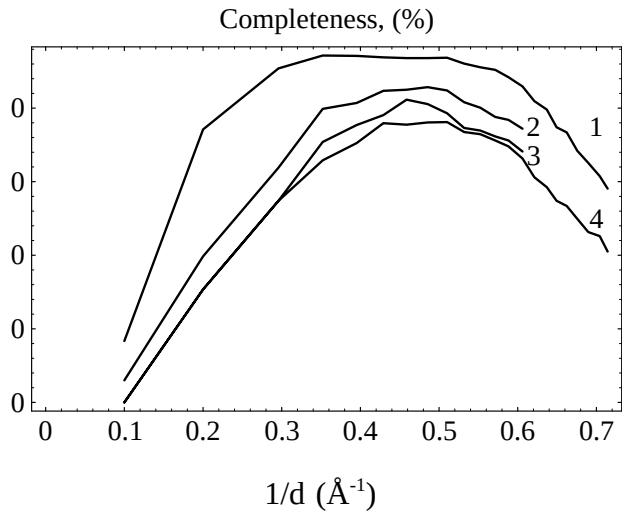

Fig. 3.1. Completeness of Laue data versus resolution. 1 denotes all data in BLD, 2 - all data in SLD, 3 - single reflections in SLD, 4 - single reflections in BLD.

For further comparison, agreement factors $R_{mono} = \Sigma \mid I_o - I_{mono} \mid / \Sigma(I_o + I_{mono})$ were calculated with respect to monochromatic data (Table 3.1). Considering those reflections which are present in both data sets SLD and BLD, the improvement in the estimates is due to the different treatment of higher harmonics. It is further due to the fact that observations which introduce degenerate terms into the least-squares matrices were omitted during the standard processing, but not in the Bayesian processing. For reflections which are present in BLD only, the accuracy in the estimates is lower, since all these correspond to degenerate components.

Hajdu et al. (1991) showed that Patterson methods are very sensitive to reveal the effects of lack of completeness in data sets. We compared Patterson functions calculated with the Laue data to monochromatic Pattersons. The correlation factor was found to increase from 0.60 to 0.79 in going from SLD to BLD.

In many time-resolved Laue diffraction studies, the most important criterion will be the contrast and quality in electron density maps calculated with model phases. Figs. 3.2 - 3.3 demonstrate that the quality of electron density maps of BPT improved

Table 3.1 R-factor versus resolution from a comparison of Laue data to monochromatic data.[1)] Single reflections only. [2)] Reflections, contained both in SLD and BLD.[3)] Reflections, contained in BLD only

Resolution shell,	R_{mono}					
(Å)	SLD[1)]	SLD	BLD[1)]	BLD[2)]	BLD[3)]	BLD
5.00	0.045	0.050	0.046	0.031	0.15	0.099
3.02	0.046	0.055	0.037	0.043	0.092	0.067
2.39	0.053	0.066	0.051	0.058	0.16	0.073
2.09	0.061	0.070	0.060	0.066	0.26	0.079
1.90	0.079	0.081	0.076	0.073	0.41	0.107
1.76	0,12	0.12	0.10	0.11	0.39	0.14
1.65	0.17	0.18	0.16	0.17	0.39	0.20
1.58				0.19		0.23
1.51				0.23		0.27
1.45				0.27		0.30
1.40				0.28		0.31

dramatically when Bayesian estimates were included. In order to simulate a near-to-real Laue study, we used phases from a low temperature (223 K) structure of BPT in an aqueous solution containing 70% methanol (Viehmann & Bartunik, 1993); all solvent molecules were removed from the model. Ten cycles of isotropic refinement against F_o were carried out for both data sets with the program SHELXL93 (Sheldrik, 1993). This yielded final $R= \Sigma \mid F_o-F_c \mid / \Sigma F_c$ values of 0.23 for SLD, and 0.24 for BLD at 1.65 Å resolution (R=0.26 for BLD at 1.4 Å resolution). $(2F_o-F_c)$ difference fourier maps that were calculated with SLD showed quite low contrast; about twenty discontinuities were visible along the main chain. With BLD, the number of main-chain cuts was reduced to one. In total, the Bayesian procedure resulted in $(2F_o-F_c)$ maps of comparable quality as monochromatic maps that were calculated at the same nominal resolution.

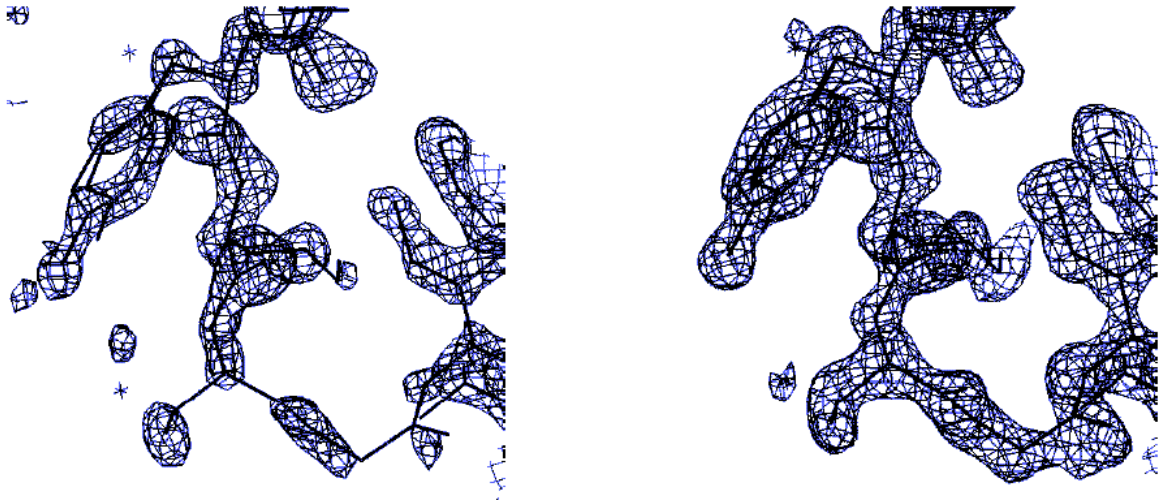

Fig. 3.2. (2Fo-Fc) electron density maps for Tyr 59, Lys 60, Ser 61 and Gly 62. (a) Structure factor amplitudes from SLD, (b) from BLD.

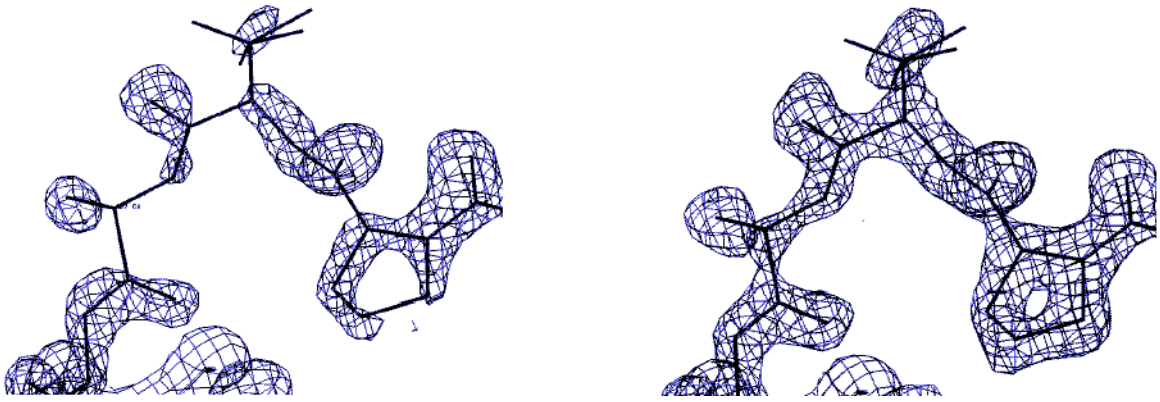

Fig. 3.3. (2Fo-Fc) electron density maps for Pro 124, Thr 125 and Ser 27. Structure factor amplitudes from SLD (left) and (b) from BLD (right).

In order to quantitatively assess the improvement in the quality of the maps, (F_o,φ_{calc}) maps that were calculated with structure factor amplitudes from both Laue data sets were analyzed in their correlation with monochromatic maps. In this case, phases were taken from an atomic resolution BPT model (Popov & Bartunik, 1996). An overall correlation factor, *c,* was calculated:

$$c = (<\rho_m \rho_L> - <\rho_m><\rho_L>)(<\rho_m^2> - <\rho_m>^2)^{-1/2}(<\rho_L^2> - <\rho_L>^2)^{-1/2},$$

ρ_m refers to the monochromatic (1.0 Å resolution), and ρ_L to Laue electron density. We obtained *c*=0.79 and 0.90 for the maps with SLD and BLD coefficients, respectively. This corresponds to real space R-factors (Bränden & Jones, 1990) of 0.34 and 0.16, respectively. Further, all atoms (protein and solvent) in the model were subdivided into bins according to the atomic B factors. The bin width was set to 5 Å^2. Correlation factors were calculated separately for areas surrounding the atoms from each bin. Fig. 3.4 shows the correlation coefficients as function of the atomic temperature factors. As one would expect, the improvement in the map quality is most evident in parts of the structure with high B values.

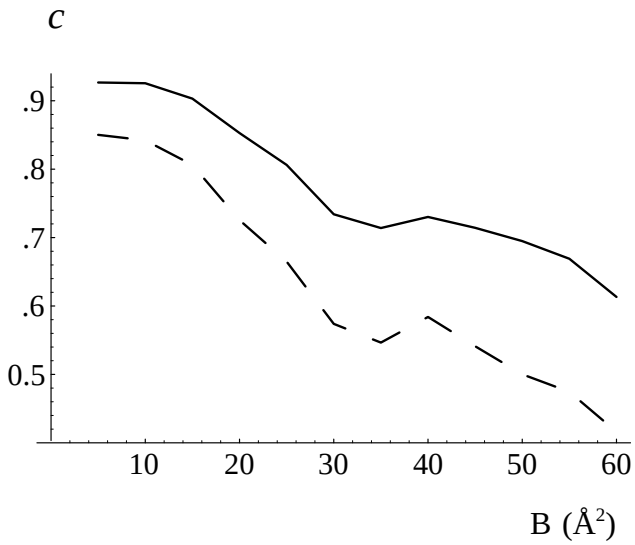

Fig. 3.4. Correlation factor derived from a comparison of Laue (F_o,φ_{calc}) density maps to a monochromatic map as a function of the atomic temperature factor. Solid line for BLD, dashed line for SLD.

3.5 Contrast in Laue density maps in case of an incomplete model

The contrast in electron density maps that are calculated with a given set of Laue data depends on the quality of the phases. In time-resolved Laue studies, phases will be known for the initial structural state. The interpretability of Laue density maps for an excited state, e.g. a reaction intermediate, will depend both on the differences in the phases between the two structures and the quality and completeness of the Laue data.

We analyzed the contrast in electron density maps that may result from Laue studies of protein structures undergoing extended conformational changes. As a hypothetical model, we considered a structural transition from bovine trypsinogen to a trypsin-like conformation. The activation domain comprising 33 residues (i.e., 15% of the molecule) is disordered in trypsinogen and ordered in trypsin (Bode & Huber, 1986). We aligned (with the program O (Jones et al., 1991)) the structural model of trypsinogen to that of trypsin; the disordered residues were omitted. The remaining residues have an r.m.s. deviation of 0.26 Å. The r.m.s. phase difference between the two models is 48^{o}. We take trypsinogen as the "initial state" and trypsin as the "final state" of the disorder-order reaction. We assume that Laue diffraction patterns were recorded from the final state, and that Bayesian processing yielded the Laue data ("BLD") as described above in the test application to the trypsin. The BLD structure factor amplitudes were combined with the phases from the initial state model. Fig. 3.5 shows an example of a $2F_o$-F_c difference Fourier map at 1.4 Å resolution covering a segment of the final state structure which is disordered in the initial state. In this "omit map", the backbone is quite well defined; in fact, it would be possible to trace the main chain for the whole activation domain despite the lacking phase information. Furthermore, most side chains (except of Gly and Ala) can be identified. This simulation demonstrates the feasibility of detecting even very extended conformational changes in Laue density maps at high resolution if the quality and

completeness of Laue structure factor amplitudes are as high as obtained in the Bayesian processing of the trypsin data.

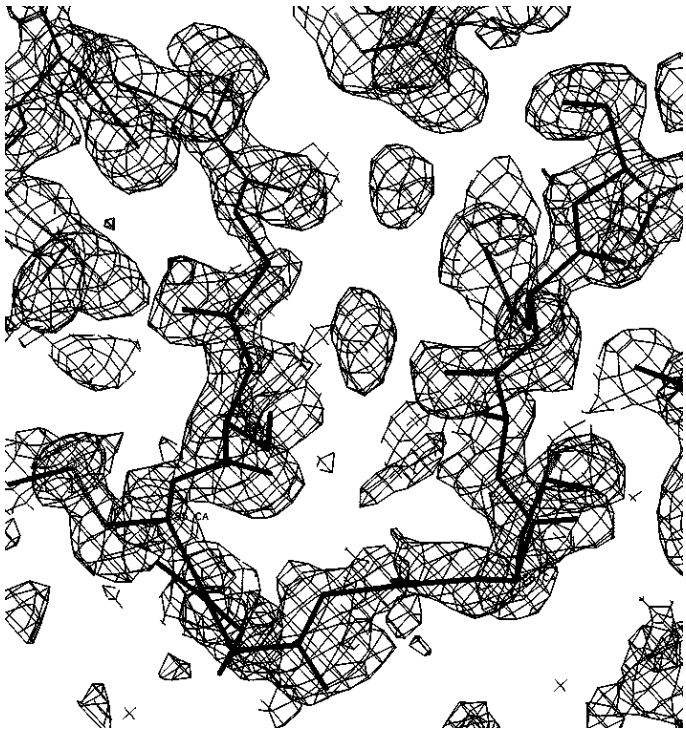

Fig. 3.5. (2Fo-Fc) electron density map calculated with (BLD) structure factor amplitudes from BPT and phases from trypsinogen for a section containing the residues 184-188. Bold lines indicate the residues that were omitted from the model.

Chapter 4
Applications of the Bayesian method in time-resolved experiments

In the following, the examples of possible applications of the Bayesian method in time-resolved experiments are discussed.

4.1 Stroboscopic time-resolved studies of CO rebinding in myoglobin

Myglobin (Mb) reversibly binds small gaseous ligands like O_2, NO and CO at its active site (the heme iron). The iron-ligand bond can be photolyzed by a visible light photon, allowing synchronous initiation of the rebinding reaction. In case of the O_2 and NO ligands, the short-lived intermediates are generated immediately after photolysis and decay back to the original liganded ground state within a few *p*s (Olson & Phillips, 1996). After spontaneous disscociation or photodissociation of the CO ligand at room temperature, the spectrum of disscociated CO molecule persists over hundredths of nanoseconds (Lim, Jackson & Anfinrud, 1997). Time-resolved spectroscopic studies (Esquerra et al., 1998, Franzen et al., 1995) and molecular dynamics simulations (Case & Karplus, 1979, Henry, 1993, Sagnella et al., 1999)

showed that the process of CO rebinding proceeds in steps that involve at least two intermediates and stuctural transitions characterized by movements of the ligand within the protein matrix and protein relaxation (Schlichting & Chu, 2000). Putative positions of the CO ligand ("docking sites") in the intermediate structure were identified by cryotrapping techniques and static X-ray crystallography at medium resolution (Schlichting et al, 1994, Teng et al., 1994, Hartmann et al., 1996, Ostermann et al., 2000, Chu et al., 2000, Brunori et al., 2000) and at ultra-high resolution by Kachalova and co-workers (2001).

The protein relaxation and the sequence of the events following ligand rebinding at physiological temperatures in theory may only be followed in time-resolved experiments. The first nanosecond time-resolved Laue experiment (Srajer et al., 1996) showed (at about 2.0 Å[3] resolution) the disappearence of electron density at the position of the bound ligand 4 *n*s after the flash photolysis and its rebinding on a *µ*s- to *m*s time scale. In this application Moffat and co-workers used a strategy that was based on the requirement (set by the deconvolution method) of repeating Laue exposures at a great number of different crystal orientations. The acceptable total X-ray dose, however, was limited by radiation damage. Therefore, each exposure corresponded to a relatively small number of X-ray photons, and hence was limited in resolution.

The results of the monochromatic studies by Kachalova et al. (1999, 2001) showed that much higher resolution and completeness is required for reliable interpretation of the small-amplitude motions. In previous studies at lower then ultra-high resolution, the geometry of the Fe-C-O unit was not correctly identified, not even in the stable MbCO structure. Furthermore, following the transition of the ligand between the different docking sites and the concerted ligand-induced conformational changes of the haem and the protein matrix, as it was possible on the basis both of stable ligation states (Kachalova et al., 1999) and of frozen structural intermediates (Kachalova et al., 2001), is feasible only on the basis of ultra-high resolution data.

[3] The data nominally extended to 1.8 Å resolution, but the completeness was <30% in the resolution range from 2.0 to 1.8 Å

In order to reach highest possible resolution and completeness in the Laue data for a given total X-ray dose, one may minimize the number of different orientations. This is possible when using the advanced Bayesian data processing methods. In this chapter, the application of the Bayesian analysis to Laue data from an MbCO crystal in the ground state is described. The data were collected as part of the development of techniques for flash-photolysis studies. We aimed to verify the experimental conditions of the time-resolved diffraction measurements, in particular to answer the question whether the quality of the data obtained from a small number of Laue exposures will be sufficient to correctly interpret the electron density of the bound ligand.

4.1.1 Data Collection and processing

An experimental setup for stroboscopic time-resoved experiments has been developed at the wiggler beamline BW6 at DORIS. It includes a mirror chopper system for extraction of individual synchrotron bunches in single- or two-bunch mode of the storage ring operation and a system for synchronization of the X-ray and laser pulses (Kosciesza and Bartunik, 1999, Kosciezsa, 2000). Each exposure was accumulated over 3000 to 10000 bunches on an IP detector in order to extend the Laue data from a Mb crystal to high resolution ≤1.7 Å; the repetition rate was ~2Hz.

Laue exposures were recorded at six different crystal orientations covering equidistantly a total range of 110 degrees of rotation. For each exposure 4000 bunches were extracted using the mirror-chopper and accumulated on the IP detector (Mar345, X-Ray research, Hamburg). A typical diffraction pattern is shown on Fig. 4.1. The crystal exhibited a strongly anysotropic mosaic spread, which is generally observed for monoclinic Mb crystals.

The Laue patterns were evaluated following the algorithm outlined in Chapter 3. The program LAUEGEN (Campbell, 1995) was used for indexing, geometric refinement and intensity integration. Reference monochromatic data from Met-Mb (1bzp PDB entry, Kachalova et al., 1999) were used for wavelength normalization. Using the Bayesian deconvolution, estimates of structure factors were obtained for

10908 independent reflections within the resolution range 17-1.7 Å. The data completeness *versus* resolution, the signal-to-noise ratio and comparison of intensities to the reference monochromatic MbCO data (1bzr, Kachalova et al., 1999) are shown in Table 4.1.

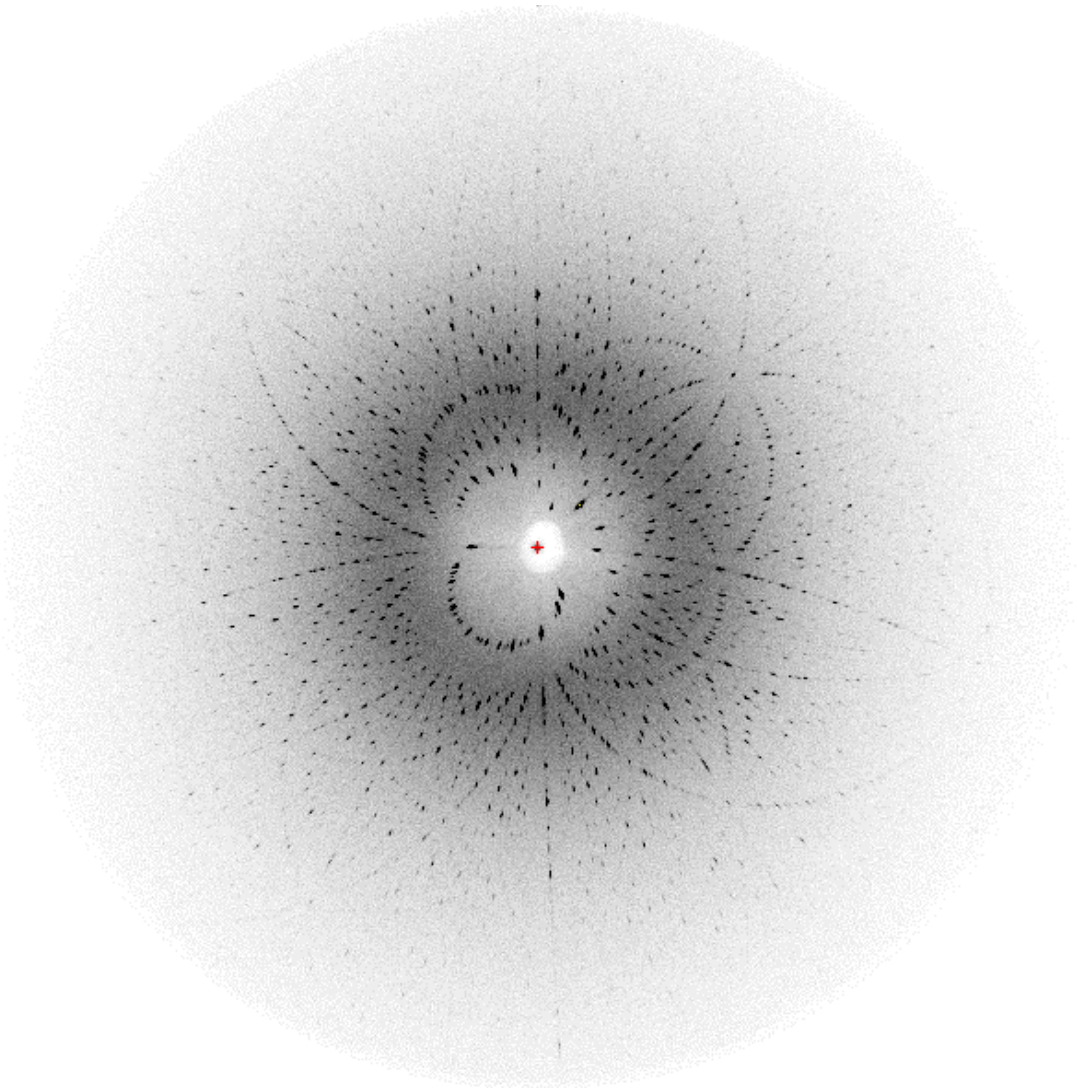

Fig. 4.1. Laue diffraction pattern of MbCO crystal. 4000 bunches (of 100 ps) were extracted with the mirror-chopper and accumulated on a Mar345 imaging plate detector, at 200 mm distance from the sample. The pattern containes contributions from 5557 reflections within the resolution range from ∞ - 1.7 Å and the wavelength range 0.60-2.0 Å.

Table 4.1 MbCO Laue data processing statistics.

Resolution Shell, Å	Completeness, % Experiment	Prediction	$\frac{\langle I_{laue} \rangle}{\langle \sigma(I_{laue}) \rangle}$	$R_{1bzr} = \frac{\sum \lvert I_{laue} - I_{1bzr} \rvert}{\sum I_{laue} + I_{1bzr}}$
17 - 10	16.4	32.9	5.4	0.07
10 - 5.5	46.6	53.4	6.9	0.14
5.5 - 3.6	62.9	66.5	8.7	0.09
3.6 - 2.4	74.1	75.9	6.9	0.11
2.4 - 2.0	76.4	77.6	5.0	0.16
2.0 - 1.8	75.9	76.9	2.7	0.28
1.8 - 1.7	75.2	76.8	2.0	0.39
17 - 1.70	73.3	75.1	5.8	0.13

The completeness of the data that was obtained from the six exposures within the resolution range 10-1.7 Å is limited only by the coverage of the reciprocal space. The incomplete coverage of the assymetric unit (75.1%) is explained by the particular orientation of the crystal, which was chosen in order to optimize the optical excitation conditions for the given crystal topology. With the crystal setting that we used, a minimum of twelve exposures would be required to cover >80% of the assymetric unit. On the other hand, a smaller number of orientations would be sufficient to reach >80% completeness if the <010> crystal face would be exposed to the laser beam. The statistical accuracy of the structure factor estimates is acceptable up to 1.7 Å resolution. Because of the low redundancy, the standard data processing procedure using least-squares harmonic deconvolution would result in much lower completeness, <20%, in the data within the range 17-2.4 Å.

4.1.2 Omit-maps of the CO ligand

An F_o-F_c difference density map was calculated using phases from the refined MbCO model (1bzr). The CO ligand was omitted from the model. The highest peak (10 r.m.s. level) in the map was found on a position were the ligand is bound perpendicular to the heme normal (Fig. 4.2). No other peaks were present on the

levels above ±4.5 r.m.s. The ligand binding geometry that was visible on the Laue difference map is close to the correct geometry as it was previously determined in the MbCO structure refined at atomic resolution (Kachalova et al., 1999).

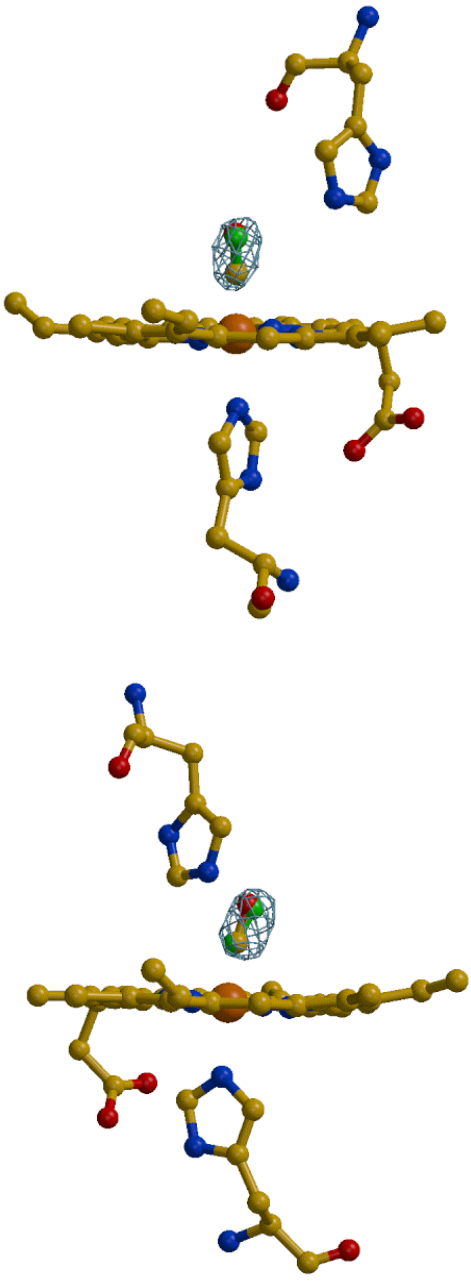

Fig. 4.2. *Fo-Fc* electron density map in the active site of the MbCO, at 1.7 Å resolution. Contouring level 5.5 r.m.s. Structure factors (*Fo*) from the Bayesian processing of Laue data. The 1bzr model of MbCO is superimposed (heme group, His64 (proxymal) and His93 (distal) residues). CO group in the position refined against Laue data is shown in green. Orthogonal views are shown.

The CO molecule was built into density and the MbCO structure refined using the program SHELXL (Sheldrick, 1997). Regarding the ligand, only the C-O bond length was restrained. The refinement converged at a crystallographic R-factor of 18%. The resulting positions of the C and O atoms shifted away from the correct positions by 0.07 and 0.17 Å, respectively. These shifts are within the range expected for an average positional error σ_x=0.22 Å, as estimated using Cruickshank's (1999) formula. In the experiment of Moffat and co-workers (Srajer et al., 1996), the resolution and the quality of the density maps were not sufficient to identify the correct orientation of the bound CO ligand.

4.1.3 Detection of reaction intermediates

This result shows that a fully occupied bound diatomic gas molecule can be identified and positioned correctly in the difference density using Laue data at 1.7 Å resolution under condition of optimum processing. In a time-resolved Laue experiment on MbCO, the intermediate states may not be fully populated. In a previous experiment (Srajer, 1996) the photolysed CO fraction was estimated to be <50%. On such a basis, it is rather unlikely that a signal-to-noise ratio similar to that what we achieved in the test experiment (4.1.2) for the ground state structure will permit to interprete the density of non-bound (intermediate) states unambiguously. For possible improvements of the conditions we note that the resolution of the MbCO data obtained with the Bayesian method was essentially defined by the Poisson statistics of the measured intensities. No effects of X-radiation damage were noticeable in the experiment described. When aiming to measure Laue data from MbCO to 1.5 Å resolution (instead of 1.7 Å), the number of bunches or the number of photons per bunch will have to be higher by a factor of 4. The estimate is based on the ratio of the average F^2 values at 1.7 and 1.5 Å ($\langle I \rangle_{1.7Å} / \langle I \rangle_{1.5\ Å} \approx 2$, with an overall temperature factor (Wilson, 1949) $\langle B \rangle \approx 12$ Å^2) for the ground state MbCO structure. The statistical error in the measured intensity is assumed to be defined by that of the incoherent background scatter. The assumption is generally valid for protein X-ray

diffraction data at the diffraction limit (Popov, 2000). It is realistic to expect that this resolution may be reached using the BW6/DORIS setup following an experimental strategy that aimes to obtain sufficient completeness in the data from a minimum number of Laue frames. The problem that remains is to find suitable conditions with respect to the laser beam parameters and the sample environment, under which a detectable fraction of MbCO molecules in a crystal can be excited repeatedly over a total period of several hours.

The long-range conformational changes in the dynamical protein domain (E and F helicies), as found from comparison of the unliganded (deoxy-) and CO- forms of Mb (Kachalova et al., 1999), involve displacements of the atoms by 0.3-0.4 Å. These amplitudes are so small that ideally ultra-high resolution (<1.2 Å) would be desired for time-resolved studies. However, in white-beam Laue mode, this would require a ~50 fold increase in the X-ray dose leading to severe radiation damage. Moreover, assuming the quality of the MbCO crystals on the same level as in the experiment described, the problem of spatial overlaps in white beam Laue mode becomes critical. As a solution to these problems, a Laue method with reduced bandwidth, $\Delta\lambda/\lambda \cong$ 1-10%, may be applied. The development and test applications of a corresponding limited-bandwidth Laue method are described in Chapter 5.

4.2 Single-shot Laue experiments: Application to cubic *2Mn*-catalase

The Bayesian method of Laue data processing does not require redundancy in the measurement of the structure factor amplitudes. This is of particular interest for structural studies of non-cyclic reactions. In such cases, sufficiently short time scales may only be reached if all information that is required for structural analysis is obtained in a single Laue exposure. The time needed for changing the crystal orientation and for reading out an area detector is long as compared to, e.g., the typical lifetimes (1-10 ms) of enzymatic reaction intermediates at ambient temperature. Even in the case of reactions which may be cycled (e.g., photodissociation of ligands from the iron in heme proteins), single-shot experiments

may yield better results since they will be less affected by radiation damage or by variations in the excitation conditions.

The geometric conditions for a single-shot Laue data collection were analysed by Clifton, Elder and Hajdu (1991). They estimated that sufficiently complete (>80%) sampling of the reciprocal space may be achieved from a single Laue exposure, if the crystal symmetry is high (Laue class *4/mmm* or higher). A more realistic estimate has to take into account that the actual completeness of the data is further reduced due to the energy overlapping reflections and critically depends on the deconvolution algorithm. We investigated the experimentally achievable completeness for two examples. The first example was based on our test experiment (Chapter 3), that was carried out with an orthorhombic crystal. In this case, a maximum of 55% of the asymmetric unit may be covered with a single crystal setting. In this experiment, the actual maximum coverage by a single exposure was 27%, due to the symmetric crystal setting. Most of the reflections were observed twice. Let us consider the doublets in the data that are truncated at 1.65 Å resolution. The redundancy was sufficient for least-squares deconvolution of 28 systems with four observations each. 466 doublets were represented by two observations; for these doublets the relative cross correlation was -98% to -100%, indicating complete degeneration. The standard procedure yielded a completeness of 9% (R_{mono} = 0.065) at d > 3.3Å, and of 18.1% (R_{mono}=0.091) at 3.3 Å < d < 1.65 Å. The Bayesian processing up to the geometrical resolution limit yielded much higher completeness, namely 24.8% of the unique reflections (i.e., 90.5% of the recorded reflections) at d > 3.3 Å, and 21.3% (85.2%) at 3.3 Å < d < 1.65 Å. R_{mono} was 0.125 at d > 3.3Å, and 0.108 at 3.3 Å < d < 1.65 Å.

As a second example, we used a cubic structure, 2Mn catalase from *Thermus termophilus* (TTC, Antonyuk et al., 2000), in order to investigate the possibility to obtain an interpretable electron density map from an application of the Bayesian method to a single Laue exposure. The crystals[4] belong to the space group $P2_13$, with the cell constant a=133.4 Å. The structure contains 604 amino acid residues (two identical subunits) in the assymetric unit. The mosaic spread of the cubic catalase crystals was extremely small, $\mu < 0.05^{o}$.

A Laue exposure from a TTC crystal was recorded on the beamline BW6 at DORIS using the MAR imaging plate detector (diameter 300 mm). The wavelength band was 0.52 - 2.1 Å. The source spectrum was measured using the Laue exposure from a reference crystal (trypsin). It was essentially identical to that shown in figure 2.4.

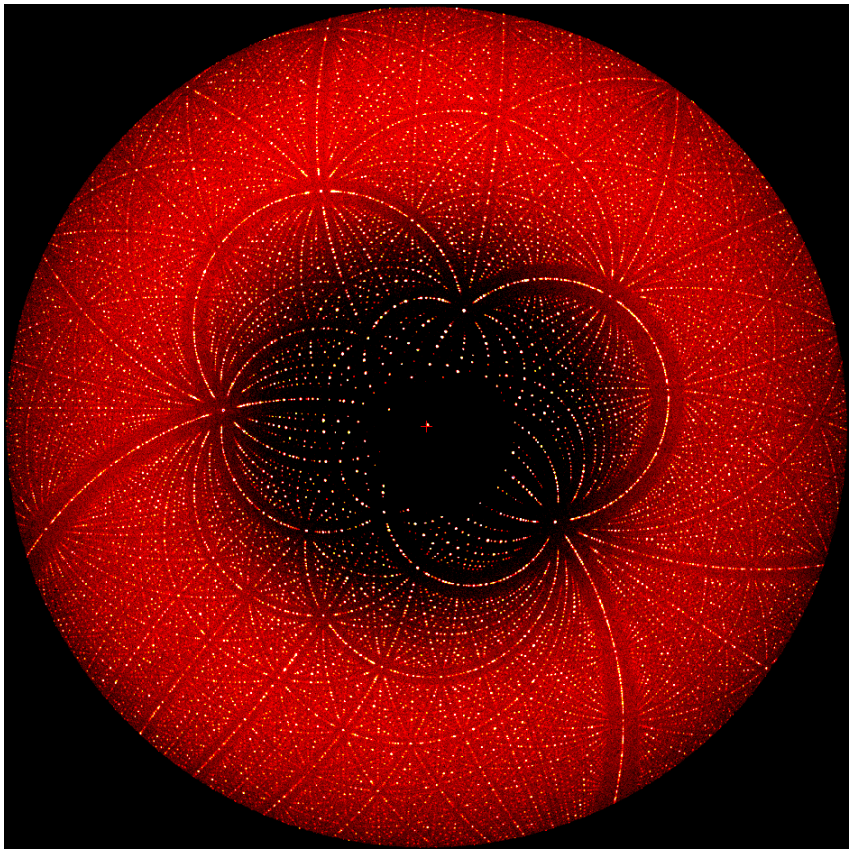

Fig. 4.3. Laue diffraction pattern of cubic *2Mn*-Catalase.

In order to spatialy separate the reflection spots, the crystal-to-detector distance had to be rather long, 400 mm. The Laue pattern (Fig. 4.3) contained contributions from 65000 reflections to the geometrical resolution limit of 1.45 Å. With the Bayesian method structure factor amplitudes for 17500 independent reflection were obtained to 2.5 Å resolution. The completeness was 69% in the resolution shell 20 - 5 Å and 78% in a

[4] The crystal was kindly donated by V. Barynin, Sheffield University, UK.

shell 5-3.5Å. The completeness dropped to 60% at high resolution (3.5-2.5 Å), due to the limited size of the detector. The resulting electron density maps were of good quality. Sections of the map are shown in Fig. 4.4.

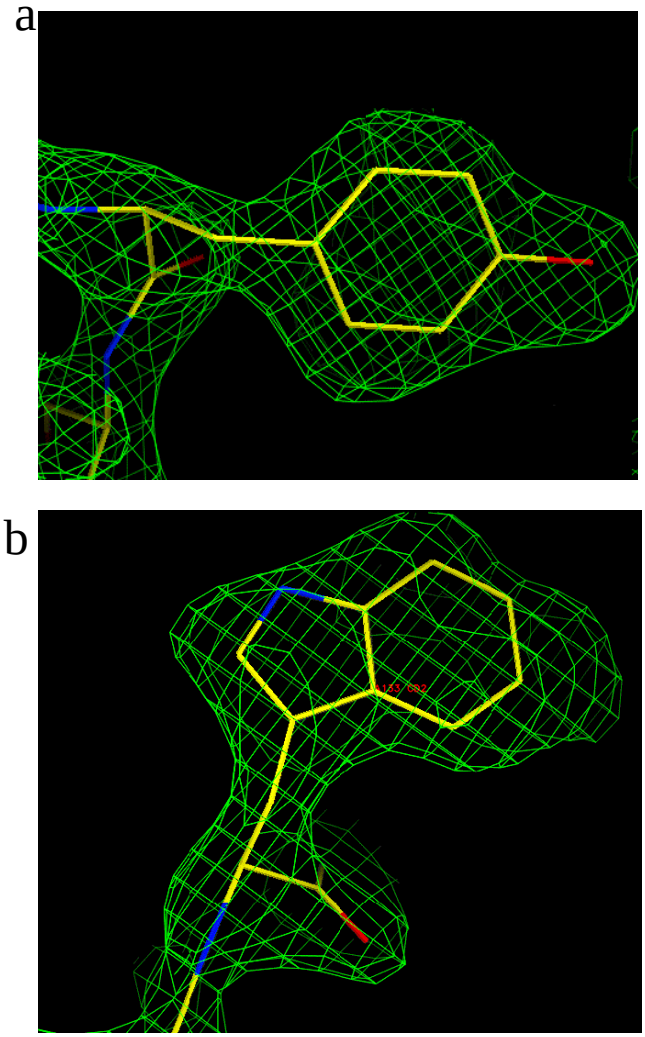

Fig. 4.4. Electron density maps of TTC calculated at 2.5 Å resolution using the structure factors that were extracted with the Bayesian method from a single Laue exposure. MIR phases after 2-fold non-crystallographic symmetry averaging (Barynin, 1996) were used for map calculation. (a) Tyr 166, (b) Trp 133.

Chapter 5
Limited-Bandwidth Laue Method

The effect of the crystal mosaicity on the white-beam Laue diffraction patterns has been discussed in Chapter 1. In many cases, the mosaic spread is too broad for high high-resolution data collection. In the example of Ras-P21 in the presence of caged ATP (Schlichting et al., 1990) white-beam Laue data analysis was limited to 3.5 Å resolution, whereas conventional monochromatic data could be measured to ultra-high reresolution close to 1.0 Å (Scheidig, 2000). Unless area detectors with sufficient energy resolution become available, the only solution to the mosaicity problem lies in a reduction of the wavelength bandwidth in the incident beam. The maximum bandwidth is defined by the mosaic spread. For X-rays with ~1 Å wavelength, the wavelengths spread within a Laue reflection from a typical protein crystal with 0.5° mosaicity[5] varies from 5% at 3 Å to 1.5% at 1.0 Å resolution. A bandwidth that matches that of the reflection from a protein crystal may be produced by an undulator or by specific X-ray optics.

Undulator sources. In a pioneering study at a CHESS undulator (Szebenyi et al., 1992), 120 ps single bunch diffraction patterns of hen egg white lysozyme (HEWL), extending to 3 Å resolution, were measured and evaluated. The single-line 162-pole

[5] Full width of the reflection.

undulator at the ESRF (U20, Wulf et al., 1997) delivered a bandwidth of 5% with integrated photon flux of the order of 10^{17} photon s^{-1} mm^{-2}. Test Laue experiments from this instrument were reported (Bourgeois, Wagner and Wulff, 2000).

Multilayer monochromators. For example, silicon-tungsten multilayers (Osmic Inc.) with 24 Å spacing deliver <2% bandwidth with 50% reflectivity. This type of monochromators was also used in earlier tests of synchrotron Laue crystallography at CHESS (Moffat et al., 1985). Later, application of multilayer optics was suggested for use in combination with the rotation method (Deacon et al., 1998). Due to a 40-fold gain in the photon flux, fast rotation data collection became possible on a bending magnet beamline. However, the data quality and resolution were limited, mostly due to the complicated spot shapes and spatial overlaps. Such a problem may be avoided by measuring still exposures. Application of the multilayer monochromators on insertion device beamlines may be complicated by the low heat conductivity of multylayer materials that may lead to loss in flux density or even to damage of the multilayer structure.

Graphite monochromators deliver a bandwidth of 2-7%. The thermal properties of graphite permits to use such monochromators even in a focused white wiggler beam without cooling. This makes graphite a convenient tool for simulating undulator conditions. In this chapter, the development of the limited bandwidth Laue method using a double-crystal graphite monocromator on the wiggler beamline BW6 at DORIS is described. The specific aim of these studies was to verify the possibility of obtaining ultra-high resolution (<1.2Å) data.

Similar methods are used in could-neutron diffractometry, employing 8-20% bandwidth at mean wavelength of ~3.5 Å (Cipriani et al., 1996). Another possible solution may be based on the use of the divergent X-ray beams (Ho et al., 1998). Applications of the method in combination with microfocusing capillary optics (Bilderback, Hoffman & Thiel, 1994) may be particularly interesting.

5.1 Experimental

5.1.1 Monochromator setup

A double-crystal graphite monochromator ([002] reflection) has been installed on the beamline BW6 after the focusing toroidal mirror (gold coated, critical wavelength 0.62 Å at 4 *m*rad reflection angle), at short (1.2 m) distance from the sample position. A short distance is required, because of the broadening of the angular divergence that is caused by the C[002] monochromator in the direction perpendicular to the scattering plane. The beam divergence is defined by two X-Y slit systems, positioned at 0.1 m and 1 m, respectively, from the sample. The energy profile of the beam was measured at the sample position by scanning a silicon crystal (Fig. 5.1). The bandwidth was 2.5% (FWHM). The monochromator crystals that we used (Union Carbide, Inc.) had a severe blocking structure (domains), which explains the observed asymmetry in the energy profile. Contamination with monochromator harmonics can be avoided by choosing a suitable cutoff wavelength of the X-ray pre-mirror. At the same time, no discontinuities in the spectrum due to absorption by the gold coating of the two mirrors should be present. Optimum conditions were obtained at a wavelength $<\lambda>$ = 1.072. The integral photon flux through a 0.3x0.3 mm^2 aperture, as calculated from ionization current measurements, was on the order of 10^{12} photon/sec, i.e. by a factor of 30 higher as compared to the standard monochromatic conditions at BW6 (double-crystal Si[111] monochromator, 0.05% bandwidth (Blume et al., 2001)).

5.1.2 HEWL high-resolution data collection and processing

Tetragonal crystals of hen egg white lysozyme (HEWL) were used for test data collection at 100K with a MarCCD detector (2K, 130 mm diameter, 0.065 mm pixel, 8 sec. duty cycle). For crystals with outer dimensions of 0.2x0.2x0.3 mm, an exposure time of 1 s was sufficient to observe diffraction spots beyond 1.2 Å resolution. Examples of the diffraction patterns are shown on Fig. 5.2. As measured separately under standard monochromatic conditions, the mosaicity of the crystal was

μ=0.2° (FWHM)[6]. Considerable elongations of diffraction spots are visible in the Laue pattern. Nevertheless, only a small percentage (23% at 45mm detector distance, 16% at 65 mm) of the reflections are spatially overlapping. Due to the decrease in the reflection bandwidth with increasing scattering angle, the decrease in the average reflection intensities between low- and high-resolution reflections is much steeper as compared to that in monochromatic rotation exposures.

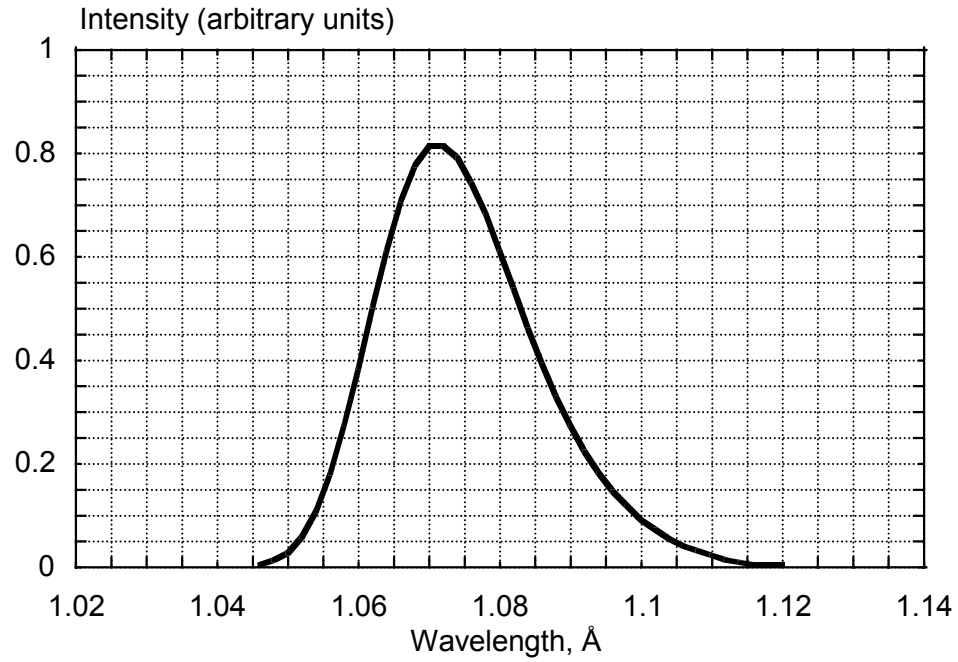

Fig. 5.1. Incident beam intensity distribution versus wavelength, obtained from the double-crystal graphite monochromator on BW6.

[6] corresponds to ~0.5° full width of reflection.

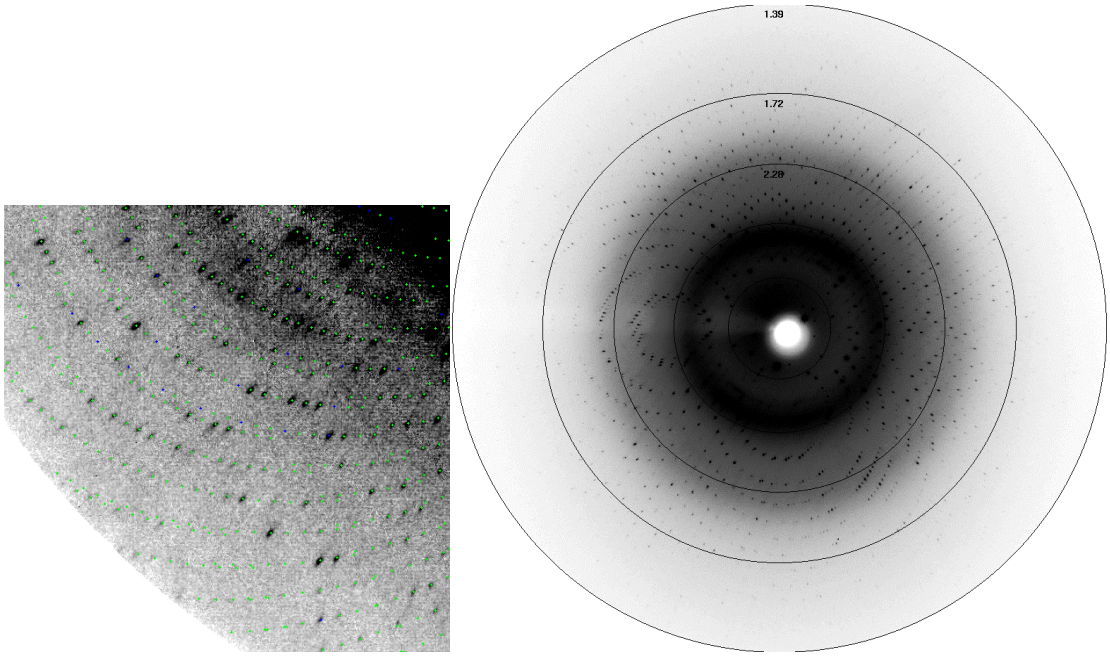

Fig. 5.2. 2.5%-bandwidth Laue diffraction patterns of tetragonal lysozyme measured on a 130mm MarCCD detector, at 45 mm (left, zoomed fraction of the diffraction pattern) and 60 mm (right) distances. Mean wavelength $\langle\lambda\rangle$=1.072 Å. Resolution at the edge of the detector is 1.18 Å (left) and 1.4 Å (right). Green spots indicate the predicted reflection positions.

We used 1° angular intervals between frames for high-resolution (1.18 Å) measurements. Fig. 5.3 shows the resolution dependence of the angular range in reciprocal space that is covered by a single exposure in the plane normal to the rotation axis, for intensity levels corresponding to 50, 10 and 5%, respectively, of the maximum intensity. Smaller intervals, 0.5°, were used for collecting low-resolution data (>1.5 Å). In total, 50 frames using the full beam intensity and 100 frames with 20-fold attenuated beam were measured, with 1 s exposure each. The diffraction images could be indexed without any difficulties. We used the program DENZO (Otwinowsky and Minor, 1997), treating the images as pseudo-rotation exposures. Alternatively, the images may be indexed as Laue patterns using the program LAIP that we developed (Chapter 6). The refinement of the orientation parameters and intensity integration were done using the program LAUEGEN (Campbell, 1995), employing profile fitting

with variable radial elliptical masks and deconvolution of spatially overlapping reflections.

When scaling the data, we realized that the wavelength spectrum drifted by 0.02 Å towards shorter wavelengths during the data collection, apparently due to the thermal expansion of the first monochromator crystal caused by heating in the white beam. In order to correct for this effect, the wavelength normalization curve was determined separately for each data frame by a binning method using redundant measurements at different wavelengths (see Chapter 2.3.2). From each curve the shift in the mean wavelength was determined as a function of the frame number. For a final scaling, the reference normalization curve (Fig. 5.1) was used after applying the corresponding shifts. No significant variation in the monochromator bandwidth, hence in the shape of the normalization curve, was detected. The $|F|^2$ values from all frames were scaled together and merged using SCALEPACK (Otwinowsky and Minor, 1997). Effects of radiation damage were noticeable, resulting in a drop in the scattered intensity by approximately a factor of two towards the end of the data collection. The statistics of the final results of the data processing are given in Table 5.1.

The finite bandwidth of Laue reflections was not taken into account in the wavelength normalization procedure. The magnitude of the systematic errors arising from such a treatment may be approximated by the equation (1.6), representing a relative error in the |F| as a function of wavelength and scattering angle. For the case of the small wavelength range in the incident beam, the equation may be approximated as follows:

$$\delta = \frac{\Delta I}{I} \cong \frac{1}{2} g'' \mu^2 \cot^2 \Theta \cong g'' \mu^2 (d^2 - < \lambda >^2 / 2) . \qquad (5.1)$$

Neglecting an asymmetry in the energy profile of the monochromator, we can write:

$$\delta \approx -\left(\frac{\mu}{\Delta \lambda} \right)^2 (d^2 - < \lambda >^2 / 2) \qquad (5.2).$$

In our experiment, $(\mu/\Delta\lambda)^2 = 0.02$ Å^{-2}. Thus, at high resolution (e.g. $d \leq 2$ Å), the

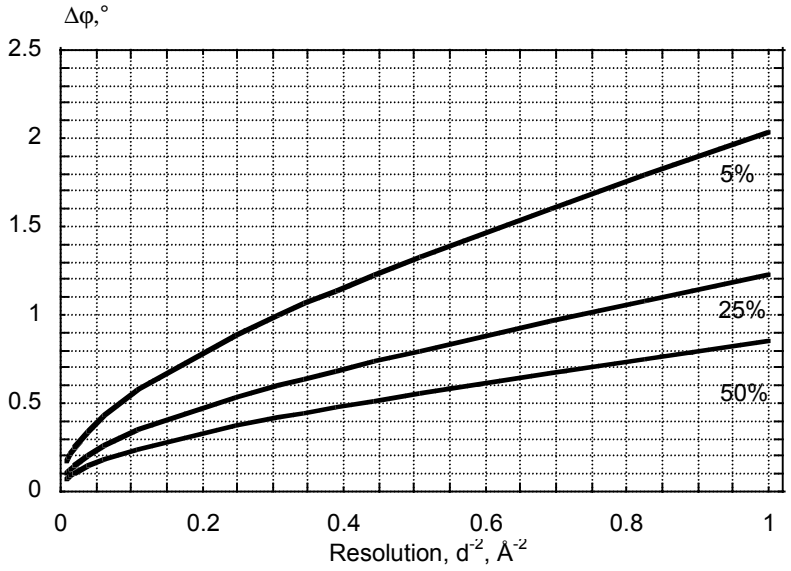

Fig. 5.3. The angular interval in the reciprocal space covered in a single exposure in the plane normal to the rotation axis as a function of resolution. Intervals corresponding to intensity levels of 50, 25 and 5% of the maximum of the energy distribution (Fig. 5.1) are shown.

systematic errors are negligible as compared to the statistical ones. However, the errors become very significant for low resolution (e.g. 5 Å $\leq d$) reflections. According to equation 5.2, the errors of this type mainly cause downscaling of low resolution reflections by a smooth function. This should not lead to severe bias in the structure refinement and calculation of electron density maps. Additional noise is visible in the merged data (especially at low resolution, see Table 5.1), which arises from an asymmetry (deviation from normal distribution) in the resolution functions of both the monochromator and the protein crystal. In order to correct for this noise, an explicit convolution of the two optical functions would have to be considered. The energy profile and divergence of the incident beam can be measured experimentally. A suitable analytical approximation would have to be found for a description of the

generally broad, anisotropic and non-uniform distribution of the mosaic blocks in protein crystals.

Table 5.1. HEWL limited-bandwidth Laue data statistics

Resolution shell, Å	Completenss, %	$\frac{\langle I \rangle}{\langle \sigma(I) \rangle}$	Redundancy	$R = \frac{\sum \lvert F_{obs} - F_{calc} \rvert}{\sum F_{calc}}$
12.0-6.05	91.0	10.0	3.6	0.162
6.05-2.32	99.6	21.7	7.5	0.110
2.32-1.91	99.6	25.9	11	0.104
1.91-1.68	99.8	22.6	13	0.101
1.68-1.54	99.5	18.4	14	0.108
1.54-1.43	99.7	13.8	14	0.116
1.43-1.35	97.5	8.7	6.9	0.131
1.35-1.29	97.4	6.9	7.1	0.143
1.29-1.23	97.3	5.1	7.3	0.160
1.23-1.18	97.0	3.8	4.7	0.205
12-1.18	98.1	17.0	12.1	0.133

The structure of lysozyme has been refined against this Laue data set at 1.18 Å resolution, starting from the 1LZ8 PDB entry (Dauter et al., 1999). The model included 1001 protein atoms, 253 water molecules, one sodium and one chloride ions. Restrained anisotropic refinement with SHELXL (Sheldrick & Schneider, 1997) resulted in a crystallographic R-factor of 13%, without any manual remodelling. The section of the 2Fo-Fc electron density maps shown in Fig. 5.4 demonstrates the high quality and the resolution of the data. An analysis of the residual density revealed the cleavage of the Cys76-Cys94 disulfide bond (Fig. 5.5) due to specific radiation damage. The result is consistent with that observed in lysozyme crystals at 100K on a third generation source (Weik et al., 2000, Ravelly & McSweeny, 2000) in experiments involving X-ray doses of the order of 10^6-10^7 Gy.

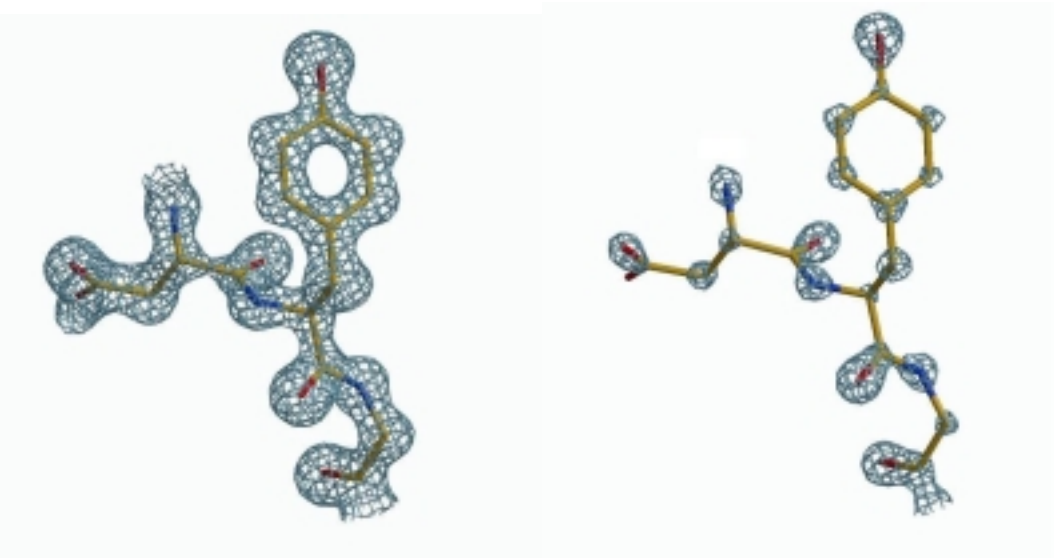

Fig. 5.4. Section of the 2Fo-Fc electron density maps around Tyr55 residue, contoured at 1.7 (left) and 5 (right) r.m.s. level**.**

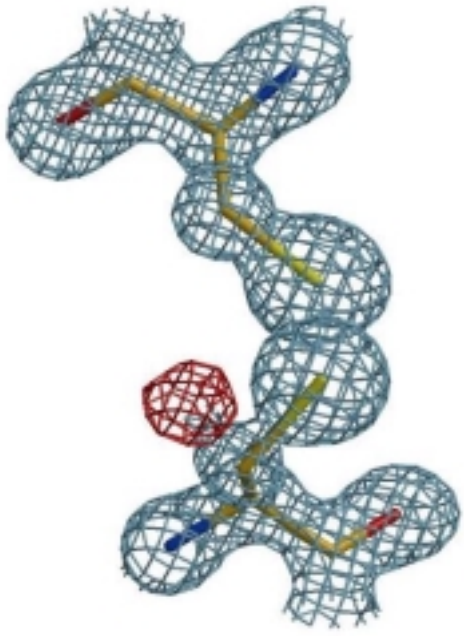

Fig. 5.5. 2Fo-Fc (blue, 1.7 r.m.s. level) and Fo-Fc (red,. 8 r.m.s. level) electron density maps of Cys76-Cys94 disulfide bond. The difference density indicates a location of the sulfur atom in an alternate rotamer conformation corresponding to an S-S bond that has been cleaved due to specific radiation damage.

Chapter 6

Automatic Indexing of Laue Patterns

Indexing is a necessary stage involved in processing diffraction data. In monochromatic methods, the term "indexing" usually assumes determination of the crystal cell and its orientation. The indexing methods (e.g., Vriend & Rossman, 1987, Kabsch, 1988) are based directly or implicitly on the Patterson transformation of the diffraction pattern, which is equivalent to the Patterson function of the crystal (or its special projection). Unlike monochromatic conditions, in case of the white-beam Laue method the solution of the problem is hindered by the uncertainty in the wavelengths corresponding to the reflections. As a result, one can determine only the directions of the experimental diffraction vectors, but not their moduli. Carr, Cruickshank and Harding (1992) proposed a method based on visual recognition of the triangles in the gnomonic projection. It enables the determination of relative cell parameters and orientation from Laue patterns. Ravelly et al. (1996) further developed the method, basing it on a semiautomatic procedure of plane recognition in the reciprocal space. Using sharp features in the white spectrum (like the Mo emission line of the X-ray generator) it was possible to determine the absolute cell volume in case of small organometallic compounds.

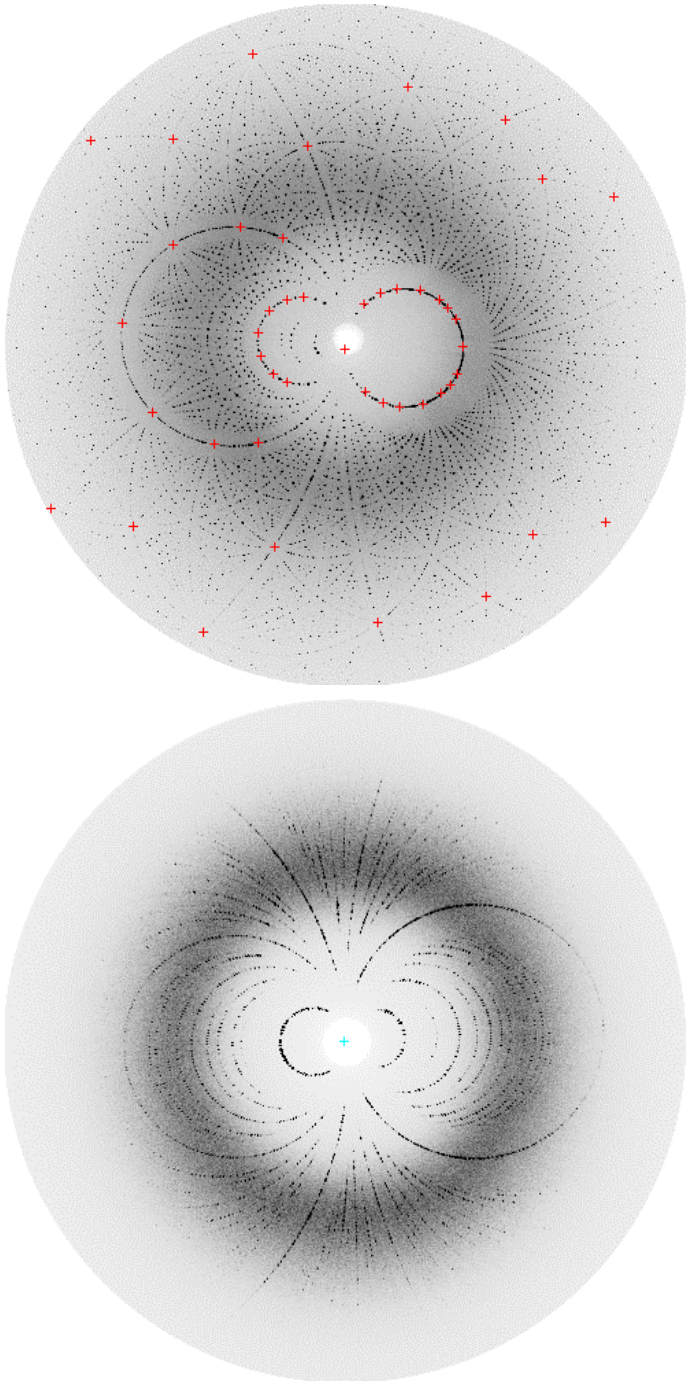

Fig 6.1. Laue patterns of BPT (low density form, top) and yeast hexokinase p152k crystals. On the BPT Laue pattern, the nodal spots, h,k,l <3 (marked by crosses) may be found visually. This is not true in case of the hexokinase Laue pattern. The experimental details are given in Table 6.1.

Up to now a number of programs have been published for automatically determining the orientation of crystals by the Laue method. These methods require that the cell parameters are known (Ploc, 1978, Riquet & Bonnet, 1979, Laugier & Filhol, 1983, Hart & Rietman, 1982, Shulakov, 1984, Ohba, Yamada, & Watanabe, 1984, Fewster, 1984, Elder, 1986, Sheremet'ev & Lyuttsau, 1986). In general, this is the case in time-resolved studies of protein structures, were the ground state structure is usually known prior to the experiment. However, the performance and reliability of the algorithms used in these approaches are severely limited and hence not adequate under a number of important experimental conditions. The programs quoted above implement autoindexing by a trial-and-error method that is based on a comparison of intravector angles of the crystal reciprocal space with the angles between the experimental diffraction vectors. The short wavelength limit (λ_{min}) of the range used, or a guess for the resolution limit (d_{min}), specifies the upper limit of the lengths of the reciprocal lattice vectors, that can correspond to each of the reference experimental vectors selected for indexing. Both the analysis of all possible combinations when indexing reference reflections and the verification of the solution by indexing the whole experimental data set involve a large body of computations. In the case of large crystal cell parameters and/or high experimental resolution a solution may not be feasible. The complexity of the problem can be characterized by the parameter M, the maximum index value assigned to reference reflections. The available programs (e.g., LAUEGEN, Campbell (1995)) allow to obtain the result in a reasonable time (minutes) only if the reflections to be indexed have indices that do not exceed the specified limit $h^2+k^2+l^2 \leq 40$, i.e. M<4. Otherwise, the orientation is not determined. Therefore, it was suggested that reflections with small indices ("nodal" spots) should be chosen as reference reflections on the basis of visual examination of a diffraction pattern. In case of Laue patterns of crystals with large unit cell dimensions (Fig. 6.1), however, this can lead to errors or even prove to be impossible.

I suggest a new method that retains the approach of the "angular test", but drastically reduces the extent of computations without introducing any artificial limitations. The main ideas of the new approach are as follows:

i. Experimental diffraction vectors that provide the minimum amount of computations are chosen as a reference set. The vector lengths lie within the limits that are defined by the experimental resolution.

ii. All possible sets of reflection indices that match the set of angles between the reference vectors are calculated on the basis of a geometric construction, instead of searching for such sets within a pre-calculated list of all interplanar angles of the crystal.

iii. The orientation matrix is found on the basis of the reference vector set and verified by indexing the whole data set. A fast check is based on the condition of indices being integers within the wavelength range that is defined by the smallest nonzero reflection index.

In the following, the new method is described in details and applied to test cases.

6.1 Description of the method

Let the reference vectors be the experimental vectors $\boldsymbol{a}_0$ and $\boldsymbol{a}_1$ observed at Bragg angles Θ_0 and Θ_1 respectively. The indices of any vector of a reciprocal lattice from a symmetrically nonequivalent area within the sphere of radius $H_0^{max} = min(2\sin\Theta_0/\lambda_{min}, 1/d_{min})$ can be assigned to the $\boldsymbol{a}_0$ vector. We calculate the list of such indices and eliminate all the vectors that are multiples of shorter vectors.

Let the $\boldsymbol{a}_0$ vector be assigned to the indices $h_0k_0l_0$. All possible indices for $\boldsymbol{a}_1$ correspond to the reciprocal-lattice points within the cones C_1 and C_2 (Fig. 6.2) with a common axis $h_0k_0l_0$ and opening angles $\alpha_1 = \alpha_0 - \delta\alpha_0$, and $\alpha_2 = \alpha_0 + \delta a_0$. α_0 is the angle between $\boldsymbol{a}_0$ and $\boldsymbol{a}_1$, $\delta\alpha_0$ is the error in measuring angles. In addition, these nodes must satisfy the condition

$$H \le H_1^{max} \qquad (6.1)$$

We search for these nodes between the intersections of cones with layer planes that are normal to the unit-cell basis vector nearest to $h_0k_0l_0$ in the direct space. The C_1 and C_2 intersections with a layer plane have the shape of ellipses if

$$\chi + \delta\alpha_0 < |\pi/2 - \alpha_0|. \qquad (6.2)$$

Here χ is the angle between $h_0k_0l_0$ and the nearest unit-cell vector. The fact that the conic sections are ellipses essentially simplifies the further computations. Hence, we select $\boldsymbol{a_0}$ and $\boldsymbol{a_1}$ in such a way that condition (6.2) is satisfied for all possible $h_0k_0l_0$ from the calculated list.

The number of reciprocal lattice vectors assigned to $\boldsymbol{a_0}$ is determined by the volume of the above-mentioned asymmetric unit. The number of vectors assigned to $\boldsymbol{a_1}$ correspond to a difference in volumes of spherical sectors that are cut by C_1 and C_2. One can easily demonstrate that the amount of computations is proportional to $(H_0^{max}H_1^{max})^3 sin\alpha_0\delta\alpha_0$ (if $\delta\alpha_0 << \alpha_0$). The minimum of this quantity on a set of pairs of experimental vectors (taking condition (6.2) into account) is a criterion for selecting the reference pairs.

The equations that limit the search for the indices assigned to $\boldsymbol{a_1}$ are derived below. Let us consider the case when the nearest axis to the $h_0k_0l_0$ vector is the *c*-axis of a direct cell and $\alpha_2 < \pi/2$. For other cases involving the use of sections with layer planes normal to the a- or b-axes, one should redefine the cell parameters *a*, *b*, *c*, α, β, γ, and the indices *hkl* appropriately and substitute $\boldsymbol{a_1}$ by $-\boldsymbol{a_1}$ if $\alpha_2 > \pi/2$.

Let us introduce an orthogonal Cartesian coordinate system (*XYZ*) in the reciprocal space of the crystal and matrix $\boldsymbol{B}$:

$$\mathbf{B} = \begin{pmatrix} a^* & b^*\cos\gamma^* & c^*\cos\beta^* \\ 0 & b^*\sin\gamma^* & -c^*\cos\beta^*\sin\alpha \\ 0 & 0 & 1/c \end{pmatrix}, \qquad (6.3)$$

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \mathbf{B}\begin{pmatrix} h \\ k \\ l \end{pmatrix}. \qquad (6.4)$$

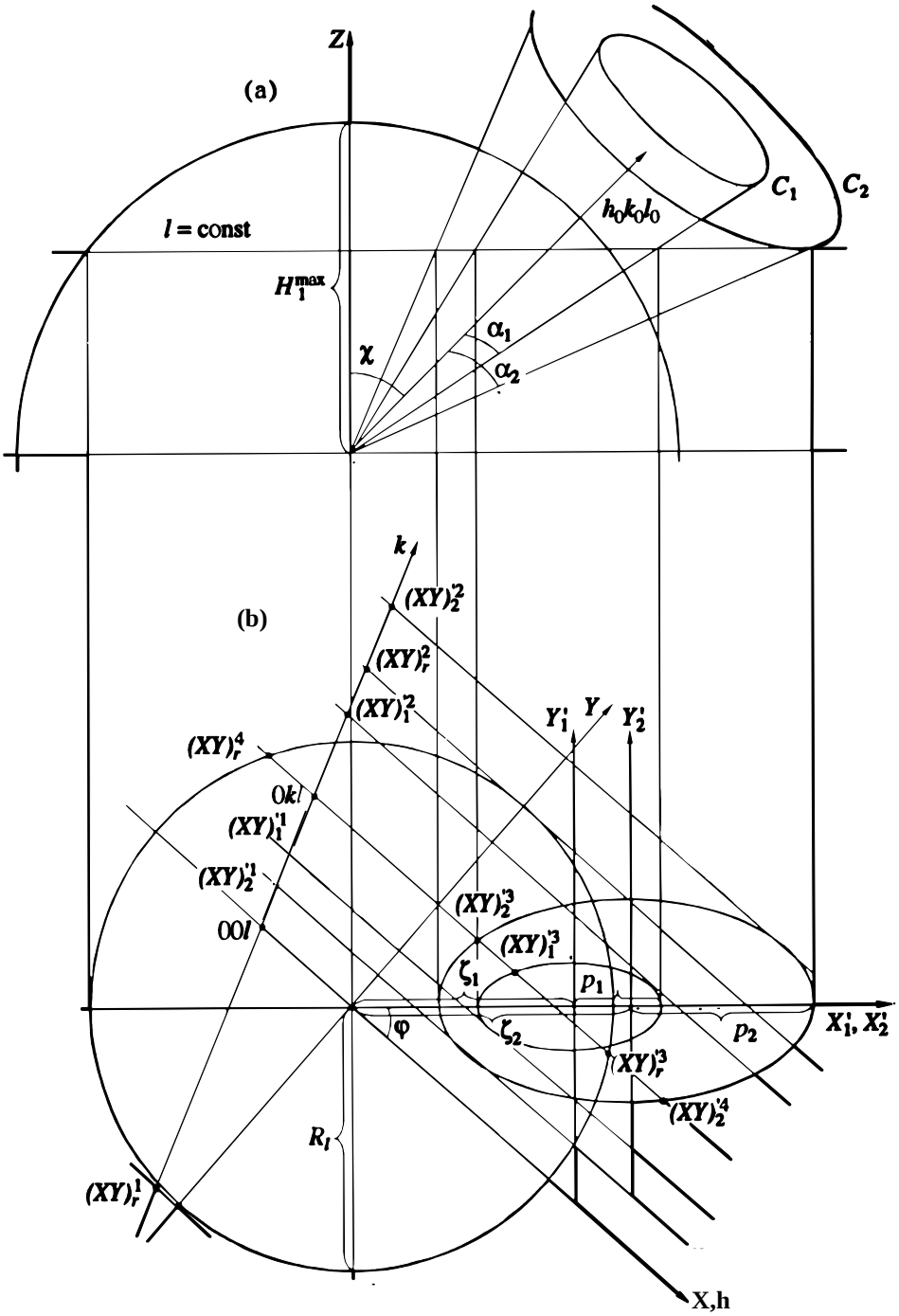

Fig. 6.2. Geometries for deriving equations for a search of the limits of the $\mathbf{a}_1$ vector indices: (a) a section of the reciprocal space by the plane passing through $h_0k_0l_0$ and Z, (b) a section by the plane with *l* = *const* .

We can then pass from reflection indices to vector coordinates (Kheiker, 1973). In this case, the *X*-axis coincides with the axis $\boldsymbol{a}^*$ of a reciprocal cell, and the *Z*-axis with the *c*-axis of a direct cell.

The highest *l*-index of the vectors is determined as ($H_1^{max} c\ S$) (see (a) in figure 6.2), where

$$S = \begin{cases} 1 & ,\alpha_1 \le \chi \le \alpha_2 \\ \cos(\alpha_1 - \chi) & ,\alpha_1 < \chi \\ \cos(\chi - \alpha_1) & ,\alpha_2 < \chi \end{cases} \qquad (6.5)$$

The limits of the exhaustive search for *h* and *k* are determined by examining the sections of C_1 and C_2 with layer planes *l* = *const*. As was noted, these sections are ellipses (see (b) on figure 6.2). It is easy to demonstrate that one of the axes of these ellipses lies on the projection of $h_0k_0l_0$ vector onto the plane under consideration. The half-lengths of the ellipse axes are

$$p_i^2 = [\ \tan(\chi - \alpha_i) - \tan(\chi + \alpha_i)\ ]^2 \frac{l^2}{2c^2}, \qquad (6.6)$$

$$q_i^2 = \left(\frac{(\xi\zeta_i + 1)^2}{(\xi^2 + 1)\cos^2\alpha_i} - \zeta_i^2 - 1\right)\frac{l^2}{c2}, \qquad (6.7)$$

where $\xi = (X_0^{\ 2} + Y_0^2)^{1/2}/Z_0$, $\zeta_i = [\tan(\chi - \alpha_i) + \mathrm{tg}(\chi + \alpha_i)]/2$, *i* = 1, 2.

We introduce the coordinate systems $(X'Y')_i$ in the plane *l* = *const*, that are related to the ellipse axes, and the matrix **W** of the corresponding coordinate transformation

$$\mathbf{W} = \begin{pmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{pmatrix} \qquad (6.8)$$

$$\begin{pmatrix} X' \\ Y' \end{pmatrix} = \mathbf{W}\begin{pmatrix} X \\ Y \end{pmatrix} + \begin{pmatrix} \frac{\varsigma_i l}{2} \\ 0 \end{pmatrix},$$

where φ is the angle between the projections of $h_0k_0l_0$ and $\mathbf{a}^*$ onto the plane under consideration. The limits for the *k*-indices of the unknown vectors are determined by the points of intersection of a nodal 0*kl* line (a straight line parallel to $\mathbf{b}^*$ and passing through 00*l* point) with the tangents to the ellipses parallel to $\mathbf{a}^*$. The equations of a tangent to the ellipse in the $(X'Y')_i$ reference system are

$$Y'_i = k_h X'_i \pm (k_h^2 p_i^2 + q_i^2)^{1/2}, \tag{6.9}$$

where $k_h = \begin{cases} Y_0 / X_0; & X_0 \neq 0 \\ 0; & X_0 = 0, Y_0 = 0 \end{cases}$,

or

$$X'_i = \pm p_i \quad at \quad X_0 = 0, Y_0 = 0. \tag{6.10}$$

The equation of a nodal line in the same reference system has the form:

$$Y'_i = k_k (X'_i - X'^l_i) + Y'^l_i \tag{6.11}$$

$$k_k = \begin{cases} \dfrac{\tan\gamma^* - Y_0 / X_0}{1 + Y_0 \tan\gamma^* / X_0}; & X_0 \neq 0, \gamma^* \neq \pi/2 \\ X_0 / Y_0; & \gamma^* = \pi/2, Y_0 = 0 \\ -\cot\gamma^*; & X_0 = 0 \end{cases}$$

or

$$X'_i = X'^l_i \quad at \quad |\gamma^* - \varphi| = \pi/2, \tag{6.12}$$

where $(XY)_i^l$ are the coordinates of a 00*l* reciprocal lattice point. Jointly solving the equations of two straight lines we obtain the coordinates of the unknown points:

$$X'^{1,2}_i = \frac{Y'^l_i - k_k X'^l_i \pm (k_h^2 p_i^2 + q_i^2)^{1/2}}{k_h - k_k}, \tag{6.13}$$

$Y'^{1,2}_i$ follows by substituting (6.13) into (6.11) for the case (6.9). If (6.10) is satisfied then

$$\begin{aligned} X'^{1,2}_i &= \pm p_i \\ Y'^{1,2}_i &= k_k(\pm p_i - X'^l_i) + Y'^l_i \end{aligned}. \tag{6.14}$$

If condition (6.12) holds, then

$$\begin{aligned} X'^{1,2}_i &= \pm X'^l_i \\ Y'^{1,2}_i &= k_h X'^l_i \pm (p_i^2 k_h^2 + q_i^2)^{1/2} \end{aligned}. \tag{6.15}$$

Performing a coordinate transformation (inverse to (6.8)) over $(X'Y')^{1,2}$ into the *XYZ* system ($Z = l/c$), and then a transformation inverse to (6.4) into the *hkl* reference system we obtain four points $K_i^{1,2}$ i = 1, 2 on the nodal line 0*kl*. K_2^1 and K_2^2 are the limits of the range of the *k*- indices of the reciprocal-lattice points that lie within C_1 and C_2 in the layer plane *l*. If $l > cH_1^{max}\cos(\chi + \alpha_2)$, one should also take condition (6.1) into account. In order to do so, we have to determine the coordinates were the tangents to the circumference of radius $R_l = [H_1^{max\,2} - (l/c)^2]^{1/2}$, which are parallel to the edge $\boldsymbol{a}^*$ of a reciprocal cell, intersect with 0*kl* nodal line. This circumference is the intersection of a sphere of radius H_1^{max} with the layer plane under consideration. In the *XYZ* coordinate system we have

$$\begin{aligned} X_r^{1,2} &= (\pm R_l - Y_l)/\tan\gamma^* + X_l \\ Y_r^{1,2} &= (X_r^{1,2} - X_l)\tan\gamma^* + Y_l \end{aligned} \tag{6.16}$$

at $\gamma^* \neq \pi/2$. ($X_l Y_l$) are the coordinates of the point 00*l* point. If $\gamma^* = \pi/2$,

$$X_r^{1,2} = 0\ ;\ Y_r^{1,2} = \pm R_l\ . \tag{6.17}$$

Applying a transformation inverse to (6.4), we pass from ($X^{1,2}_r$,$Y^{1,2}_r$, *l*/*c*)) to integer indices. Thus, we obtained additional limitations for *k* and redefine K_2^1 and (or) K_2^2 accordingly.

Furthermore, the points of intersections of the ellipses with the nodal lines possessing constant *k* and *l* determine possible *h*-indices of the required vectors. The equations of these lines in the $(X'Y')_i$ reference systems have the form

$$Y'_i - Y'_k = k_h (X'_i - X'^k_i)\ , \tag{6.18}$$

where $(X',Y')_i^k$ are the coordinates of the *0kl* reciprocal lattice point. By jointly solving the equations of straight lines we obtain the coordinates of the intersection points:

$$X'^{3,4}_i = p_i(\pm q_i(g_i - n_i^2)^{1/2} - k_h p_i n_i)/g_i\ , \tag{6.19}$$

where $g_i = (k_h^2 p_i^2 + q_i^2)^{1/2}$ and $n_i = k_h X'^k_i - Y'^k_i$. The values of $Y'^{3,4}_i$ are determined by substituting the corresponding quantities into (6.18). Otherwise, if condition (6.10) is satisfied,

$$\begin{aligned} X'^{3,4}_i &= X'^k_i \\ Y'^{3,4}_i &= q_i - (1 - (X'^k_i / p_i)^2)^{1/2}\ . \end{aligned} \tag{6.20}$$

Transformation of (6.19) or (6.20) into the system of indices defines the limits of an exhaustive search for *h*-indices. Note that for indices with $K_1^2 \le k \le K_1^1$ and $K_1^2 \le k \le K_2^2$, these limits are defined by intersections of a nodal line with C_2 only (i.e. by X_2^3, X_2^4), whereas for $K_1^1 \le k \le K_1^2$ the limits are defined by intersections with C_1 and

C_2. Considering condition (6.1), we find the intersections between this line with a circumference of radius R_l. In *XYZ* coordinates

$$\begin{aligned} X'^{3,4}_r &= \pm(R_l^2 - Y_k^2)^{1/2} \\ Y'^{3,4}_r &= Y^k \end{aligned}, \qquad (6.21)$$

Y^k is the coordinate of the 0kl point in the XYZ-reference system.

Thus, taking into account extinct reflections and reflections that exist in several orders, we determine all the indices that can correspond to $\mathbf{a_1}$, assuming that $\mathbf{a_0}$ has $h_0k_0l_0$ indices. Assigning $h_1k_1l_1$ indices to $\mathbf{a}_1$ and assuming that we know the directions of $\mathbf{a_0}$ and $\mathbf{a_1}$ in the laboratory frame of reference, we can find the crystal orientation matrix **U**. We use the algorithm proposed by Kabsch (1988) for determining the rotation matrix which provides the best coincidence between the two sets of vectors.

For verifying the solution obtained one should index all the recorded reflections. Let *X*"*Y*"*Z*" be the coordinates of some experimental vector in the laboratory reference system. Knowing the position of the reflection, the quantities (λX", λY", λZ") can be determined, whereas λ is unknown. In this case, the indices of the corresponding reflection are determined as

$$\lambda\begin{pmatrix} h \\ k \\ l \end{pmatrix} = \begin{pmatrix} \eta \\ \mu \\ \omega \end{pmatrix} = (\mathbf{UB})^{-1}\begin{pmatrix} \lambda x \\ \lambda y \\ \lambda z \end{pmatrix}. \qquad (6.22)$$

Let us denote τ = min (η,μ,ω), $\tau \neq 0$. Then, $t = \tau/\lambda$ is the smallest nonzero reflection index, $\tau/\lambda_{max} \le t \le \tau/\lambda_{min}$. A reflection considered to be indexed, if the remaining indices adopt integer values (within a certain specified deviation Δ) for some integer t from this interval and a corresponding value of λ. Let us estimate a probable error, that is, the probability of indexing a reflection with incorrect **U**. Let t take on N integer values in the interval obtained. Then, at some random value another index satisfies the criterion of integer numbers $2N\Delta$ times. The probability for two indices to simultaneously fall within Δ interval is $(2N\Delta)^2$. For the example of Δ = 0.10, incorrect **U**

will provide indexing of the reflections with N ≥ 5. Any data set will contain also reflections with small τ ($N \leq 4$). In the example of the data set where two reflections with N=1 are present, the probability of rejecting a wrong solution is 99.84%.

A necessary condition for the orientation matrix to be determined is that correct indices are assigned to the two reference vectors. The accuracy of an orientation matrix defined on the basis of two reference vectors only may be insufficient to correctly index the whole experimental set. This may occur for the structures with large unit cell dimensions. In such cases, additional reference vectors are included. When k+1 reference vectors $\mathbf{a}_i$, 0<i<k, are selected, the indices are assigned to each of $\mathbf{a}_i$ following the algorithm described above. After that, an angular cross-check is applied directly to each of the $\mathbf{a}_i\mathbf{a}_j$, $0 < j < i \leq k$ pairs. Several symmetry-related solutions will be obtained, if $\mathbf{a}_0$ lies in a symmetry plane of a crystal Laue-group or coincides with a symmetry axis.

6.2 Computer program LAIP

The algorithm described above has been implemented in a computer program, LAIP. The Laue Data Module (LDM, Campbell, 1996b) data format is used on input for the description of the experimental parameters and crystal cell. For standard diffraction geometry, the list of spot coordinates is produced by peak-search routines (e.g., *xdisp* of the HKL suite (Otwinowsky & Minor, 1997), or *marpeaks* from X-ray Research Gmbh). As an option, which is useful in case of extremely mosaic crystals, the reflections may be peaked up manually using *lauegen* (Campbell, 1995). The program then calculates the directions of diffraction vectors and their standard uncertainties. As a further option, one can input a list of the coordinates of the diffraction vectors with estimated errors. This permits to apply the program in non-standard geometry via a simple routine converting the spot positions into vector coordinates. The program outputs either an LDM data block containing crystal setting angles or an orientation matrix.

Table 6.1. Experimental conditions and results of autoindexing.

Protein	Laue Class	Unit cell, Å and °	Detector distance, mm	d_{min}	# reflections on input	M	r.m.s.d., mm
BPT	*mmm*	62 64 68	230	2.0	4617	17	0.18
Trypsinogen	3*m*1	55 109	230	2.0	3200	27	0.23
Mb	2/*m*	35 31 65 β=105	300	2.5	1568	13	0.27
Mb	*mmm*	40 49 79	350	3.0	496	11	0.46
TTC	*m*3	134	400	2.5	2819	18	0.09
HK152K	3*m*1	72 200	300	3.5	1624	22	0.24
HK-PII	*Mmm*	215 90 58	120	3.2	10	15	n.c.

The wavelength range was $0.55<\lambda<2$Å in all experiments. Autoindexing was performed on the basis of three reference vectors. The lattice type is primitive for all structures. d_{min} = resolution limits assigned for autoindexing. M = magnitude of the highest index among the indexed reflections. $<\Delta>$ = the average error in the two highest reflection indices assigned; averaging over all indexed reflections. r.m.s.d = root mean square deviation between the predicted on the basis of the autoindexing solution and measured centroids of the reflections. n.c. : not computed. BPT : bovine pancreatic trypsin, Mb : myoglobin, TTC : *Thermus thermofilus 2Mn*-catalase, HK : yeast hexokinase. HK-PII was measured using photographic film. All other data were recorded with the MAR IP detector.

The new approach of autoindexing has been applied successfully to a number of experimental Laue data sets for structures with unit cell dimensions up to more then 200 Å.

6.3 Test applications

The program LAIP has been tested with a variety of synchrotron Laue patterns from protein crystals. The data were recorded using a MAR imaging plate (X-Ray Research Gmbh) or photographic film. For two of the structures (Mb, HK PII), the crystals had broadened mosaicitsies with mosaic spread angles >0.5°. The results are summarized in Table 6.1.

The method requires rather precise knowledge of the position of the direct beam projection on the detector and the scanner offsets. These parameters can easily be determined with an accuracy ~0.2 mm, which proved to be sufficient for obtaining correct solutions in all cases.

The test applications of the method to Laue patterns of the yeast hexokinase structures are particularly interesting. For both HK-152k (Fig. 6.1) and HK-PII (Fig 6.3) cases the orientation could not be found by selecting nodal spots as a reference. The HK-152k crystal was oriented so that no nodal spot were apparent on a patterns. The Laue pattern of the HK-PII (Fig. 6.3a) was recorded 1 min after diffusion of glucose into the crystal (Bartunik et al., 1992). The process is accompanied by disordering of the crystal (Fig. 6.3 c,d). As a result, the positions of nodal spots can not be measured with the accuracy required for autoindexing. At the same time, parts of the pattern (Fig 6.3 d) shows diffraction spots that are resolved into components related to misoriented crystal blocks. Autoindexing based on these spots with indicies $h,k,l \leq 15$ was successful (Fig. 6.3b).

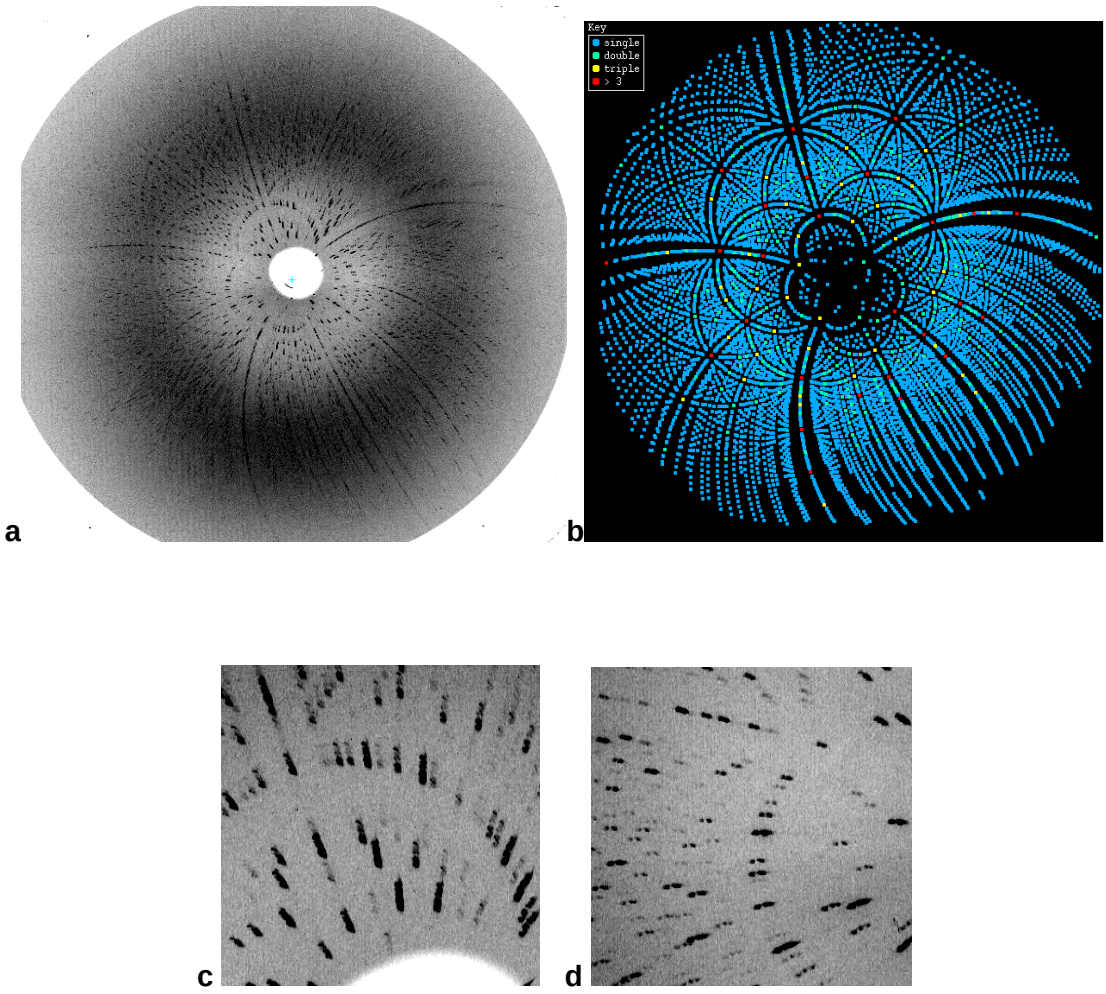

Fig. 6.3. a) Laue pattern of yeast hexokinase PII after rapid diffusion of glucose. b) Predicted pattern calculated on the basis on the results of autoindexing. c) Zoomed area which shows the disorder. d) Area of the pattern used for autoindexing.

Discussion

The Bayesian procedure of deconvoluting multiplets in Laue diffraction patterns represents a novel and very powerfull approach to solving the fundamental problems of Laue protein crystallography. As compared to all previously suggested methods of deconvolution, the Bayesian approach is the only method that provides reliable estimates of the uncertainties in the deconvoluted reflection intensity components and the resulting structure factor amplitudes. A meaningful weighting scheme may be derived from these estimates. Furthermore, the new method does not require redundant measurement of reflection intensities. Hence, Laue data collection may be limited to a minimum number of crystal orientations that are needed for a given space group in order to achieve sufficiently high completeness.

As an important consequence of the Bayesian method, no "soft" diffraction limit is needed in Laue data processing; rather, data may be evaluated to the physically relevant (wavelength dependent) diffraction limit without loss of completeness at low and medium resolution.

A test application of the Bayesian method to Laue diffraction data from orthorhombic trypsin demonstrated its power. The completeness in the Laue data set improved strongly as compared to the standard procedure followed up to now. The low-resolution hole was filled up significantly, and structure factor amplitudes were obtained up to substantially higher resolution. In fact, much higher resolution (1.4 Å)

was reached than in any previous Laue data processing. The contrast in electron density maps calculated with Laue structure factor amplitudes and model phases improved dramatically. $2F_o$-F_c difference Fourier maps became well comparable in quality to typical monochromatic high-resolution maps of the same structure.

Structure factor amplitudes of similar quality and completeness as obtained from the Bayesian processing of the trypsin Laue data may produce interpretable density maps at high resolution, even if the structure under investigation differ substantially from the structural model that provided the phases. The theoretical example of a disorder-order transition in trypsinogen, involving a r.m.s. difference in phases of nearly 50^o indicates the wide range of potential applications of Laue techniques.

In the test application to BPT, the Bayesian procedure yielded estimates for the intensity components of reflection multiplets. The accuracy of these estimates is necessarily lower than the accuracy, which can be achieved from independent measurement of (single) reflections under given white beam conditions. Nevertheless, the quality of the electron density maps improved substantially, essentially due to much higher completeness at low and medium resolution. This represents a significant gain from the Bayesian analysis as compared to the standard procedure. Most of the reflections in both data sets SLD and BLD (Chapter 3.3) at 1.65 Å resolution were represented by four observations. However, this redundancy did not yield well defined systems for least-squares deconvolution, partly due to a symmetric crystal setting. The improvement in the map quality shows that the Bayesian method successfully copes with the degenerate systems. This demonstrates the potential power of the method for applications where highest possible completeness must be obtained from a minimum number of exposures. In general, redundant observations will also be present, and the Bayesian procedure makes use of them. For example, corrections for wavelength dependent absorption must be defined from redundant observations of single reflections. Redundancy in the measurements of multiple reflections - even if they are recorded at such conditions that their wavelength normalization coefficients are linearly dependent - reduces the standard uncertainty in the deconvoluted components. The accuracy in the Bayesian estimates noticeably

increases when redundancy removes, at least partly, the degeneracy of the corresponding least-squares system. However, the existence of redundant measurements is not a necessary condition for obtaining statistically significant estimates from the Bayesian approach.

We tested the Bayesian method in applications to a well-ordered crystals. In the case of crystals exhibiting broadened mosaicity, Laue patterns contain streaky and spatially overlapping reflections. Such broadening is often observed during or short time after initiation of a reaction in the crystal (Bartunik, Bartunik & Viehmann, 1992). With the new Bayesian method there is no formal difference between the processing of partially overlapping reflections and the processing of a multiplet measured with redundancy, provided that profile fitting is applied. Our formulation extends to this case without any modification simply by considering each individual pixel as an observation and incorporating the model profile value in the coefficient a_i (eq. 2.3). In the present test applications, we took observations as uncorrelated if reflections were separated by more then one FWHM. Observations resulting from more closely spaced reflections were taken as -100% correlated, simply because the correlation coefficients were not calculated by the peak integration software of the Daresbury package. Minor changes to the existing software can lead to substantial improvements in the accuracy in the intensities of deconvoluted spatial overlaps.

The Bayesian method of Laue data processing offers new experimental possibilities in the field of fast kinetic crystallography. This refers to the studies of both cyclic single-turnover reactions and non-repeatable processes.

In combination with Bayesian data processing, *n*s time-resolved experiment on CO rebinding in Myoglobin became possible under DORIS conditions due to the fact that the number of crystal settings could be minimized. Test experiments exploiting the experimental setup on the wiggler beamline BW6 resulted in higher data quality then obtained in earlier time-resolved experiments on MbCO, which were carried out on a third generation synchrotron source. In fact, our experimental strategy and data reduction procedure, which are both based on the Bayesian method, led to superior

resolution and well-defined electron density clearly showing the correct orientation of the CO ligand.

Enzymatic reactions in crystals are practically non-cyclic. Lifetimes of the intermediates typically are in the *m*s time range. Most of the previous time-resolved experiments of this type (e.g., Hajdu et al, 1987, Schlichting et al., 1990, Stoddard et al., 1991, Duke et al., 1994) involved a total data collection time of minutes for relatively low resolution. With modern sources and detectors monochromatic methods are capable of reaching the timescale of a few minutes for full data collection. In an application to cubic TTC crystals we demonstrate for the first time that our methods allow to obtain interpretable density maps even from a single white beam shot. Thus, in the case of a sufficiently high crystal symmetry, this makes it feasible to investigate non-cyclic reactions on virtually any time scale that is permitted by the time structure and intensity of the source. At present, the time resolution that may be achieved at BW6/DORIS in such one-shot experiment is limited by the available flux density to the range of milliseconds. Considering future radiation sources, e.g. spontaneous emission of white radiation spectrum in the FEL source (Materlik & Tschentscher, 2001), a 100-*f*s long pulse will be sufficient to obtain a high resolution Laue pattern even from a multiprotein structure of the complexity of the reaction centre, provided that a suitable detector (e.g. a large-area integrating pixel array, Eikenberry et al., 1999) will be available. As long as the scattering process is faster then the deterioration of the crystalline order by radiation damage of the sample, single-shot Laue diffraction will represent a unique experimental opportunity.

The possibility to collect high-resolution data from protein crystals using a series of still exposures in a polychromatic beam with a bandwidth of a few per cent has been demonstrated using a double-crystal graphite monochromator. The data collected from lysozyme in a test experiment extended to ultra-high resolution, 1.18 Å; they were comparable in quality to the standard monochromatic data. The total exposure time required was shorter by more then one order of magnitude.

An energy bandwidth of 2.5%, as it was defined by the graphite monochromator, is especially suitable for collecting ultra-high resolution data from

small- to medium size structures (e.g., MW < 20 kDA). The total number of exposures that is required is smaller then for the monochromatic rotation method. We used 1° steps for collecting a high-resolution data set. The data set is fairly complete (>80%) in the resolution shell 1.5-1.2 Å also with a step of 3° (i.e. 17 exposures). The method is less efficient in sampling the reciprocal space at low resolution. Even with hundred exposures at very low resolution (>6 Å) a gap remains. A dispersive monochromator with a tunable bandwidth would give a possibility to overcome this problem. A multilayer with a variable number of layers, which has been suggested by Bilderback and co-authors (Deacon et al., 1998) may provide a possible solution, if the problems of heating the multilayer in the white beam may be overcome.

The scope of potential applications of the proposed limited-bandwidth Laue technique includes stroboscopic time-resolved experiments at ultra-high resolution. The method offers a number of advantages over the fine-angular interval white-beam Laue method. This is due to i) a higher signal-to-noise ratio, ii) the absence of harmonic overlaps, iii) strongly reduced effects of radiation damage, iv) the possibility to use crystals with much higher mosaicity. When using an undulator at a third generation source (e.g., U20/ESRF) 1013 photons with 2.5% energy spread may be deposited on a crystal in a single bunch exposure. Thus, 100 ps time-resolution studies at atomic resolution should be feasible already now. Considering the projected properties of the spontaneous emission radiation of the free electron laser, the time resolution will scale down to 100 fs.

Energy bandwidth of a few percent better matches the inherent bandwith of protein crystals. Thus, this technique could in principle become a better alternative to the rotation method for experiments that do not involve optimized anomalous scattering. The development of the method should, first of all, be directed towards better understanding of the optical function of protein crystals.

Conclusions

The result of the research described in this thesis is a set of new theoretical and experimental methods that permit to derive reliable structural information from Laue diffraction data. The Bayesian deconvolution may provide a breakthrough in the applicability of Laue white-beam diffraction methods which until now has mostly been limited by poor contrast and quality of the resulting electron density maps. Limited-bandwidth Laue method circumvents the problems related to the crystal mosaicity and brings the quality of structural information obtained in time-resolved mode to the level of the current standard for a static structure.

The time-resolved Laue method has not yet found wide application in structural research. This is only partly due to the limitations inherent to the diffraction method. The level of complexity in such studies – considering the choice of the system and external excitation conditions – very high as compared to static structure determination. Both stroboscopic investigations of reversible reactions on the time scales defined by the synchrotron bunch length and studies of irreversible enzymatic reactions may be set up on a limited number of systems. These systems must be suitable not only from the biochemical point of view, but also must meet the requirements with respect to the crystal quality and (possibly) even to the space group. In any case, the studies must be combined with investigations of static structural states and cryo- or chemically trapped substates at ultra-high resolution using monochromatic methods. In such combined studies, the results to expect from the time-resolved diffraction may reveal the time sequence in the population of structural substates, rather then delivering precise structural descriptions of these substates. Thus, time-resolved protein crystallography may provide valuable experimental information that is often not available from other techniques but required for a full understanding of protein structure-function relationships.

References

Andersson, I., Clifton, I.J., Edwards, S.L., F l p, V., Hadfield, A.T., Nordlund, P., Phizackerley, R.P., Soltis, S.M., Wakatsuki, S. and Hajdu, J. (1992). On the scope and limitations of the Laue metod in kinetic crystallographic studies with macromolecules. In: Time-resolved macromolecular crystallography, Eds. D.W.J.Cruickshank, J.R. Helliwell and L.N. Johnson, p. 95-102, Oxford Univ. Press, Oxford.

Amoros, J.L., Buerger, M.J. & Amoros, M.C. (1975). The Laue Method. New York: Academic Press.

Arzt, S., Campbell, J. W., Harding, M. M., Hao, Q. & Helliwell, J. R. (1999). LSCALE - the new normalization, scaling and absorption correction program in the Daresbury Laue software suite. *J. Appl. Cryst*. 32, 554-562.

Bartsch, H.H., Qichen Huang, Koelln, I., Rupp, S., Summers, L.J. & Bartunik, H.D. (1990).*Acta Cryst*. **A46**, C21.

Bartunik, H.D., Bartsch, H.H. & Huang Qichen (1992). Accuracy in Laue X-ray diffraction analysis of protein structures. *Acta Cryst*. **A48**, 180-188.

Bartunik, H.D., Bartunik, L.J. & Viehmann, H. (1992). Time-resolved X-ray diffraction studies on enzymes under cryoconditions. *Phil. Trans. R. Soc. London* **340**, 209-220.

Barynin, V. (1996). Privat communication.

Blume, J., Boesecke, P., Bourenkov, G., Kosciesza D. & Bartunik H.D. (2001). The Protein Crystallography Beamline BW6 at DORIS – Automatic Operation and High-Throuput Data Collection. *Nuclear Instr. Meth.* **A467-46,** 1358-1362.

Bilderback, D.H., Hoffman S.A., Thiel D.J. (1994). Nanometer spatial resolution achieved in hard x-ray imaging and Laue diffraction experiments. *Science*. **263** (5144), 201-203.

Bode, W. & Huber, R. (1986). Crystal structures of pancreatic serine endopeptidases. In: Molecular and Cellular Basis of Digestion, Eds. P.Desnuelle, H.Sj str m and O.Nor n, pp. 213-234, Elsevier Publ., Amsterdam.

Bourgeois, D., Wagner, U. & Wulf, M. (2000), Towards automated Laue data processing: application to the choice of optimal X-ray spectrum. *Acta Cryst*, **D56**, 973-985.

Branden, C. & Johnes, A. (1990). *Nature* 343, 687-689.

Brunori, M., Vallone, B., Cutruzzola, F., Travaglini-Allocatelli, C., Berendzen, J., Chu, K., Sweet, R.M. & Schlichting I. (2000). The role of cavities in protein dynamics: crystal structure of a photolytic intermediate of a mutant myoglobin. *Proc Natl Acad Sci USA*. **97**(5), 2058-2063.

Campbell, J W; Clifton, I J; Greenhough, T J; Hajdu, J; Harrison, S C; Liddington, R C; Shrive, A K (1990). Calcium binding sites in tomato bushy stunt virus visualized by Laue crystallography. *J. Mol. Biol*. **214** (3), 627-632.

Campbell, J.W. & Hao, Q. (1993). evaluation of reflection intensities for the components of multiple Laue diffractionspots. II. Using the wavelength-normalization curve. *Acta Cryst*. **A49**, 889-893.

Campbell, J.W. (1993). Daresbury Laue Software Suite. SERC Daresbury Laboratory, Warrington, England.

Campbell, J. W. (1995). LAUEGEN, an X-windows-based program for the processing of Laue diffraction data. *J. Appl. Cryst.* **28**, 228-236.

Campbell, J. W., Clifton, I. J., Harding, M. M. & Hao, Q. (1995). The Laue Data Module (LDM) - a software development for Laue X-ray diffraction data processing. *J. Appl. Cryst.* **28**, 635-640.

Carr, P. D., Cruickshank, D. W. J. & Harding, M. M. (1992). The determination of unit-cell parameters from Laue diffraction patterns using their gnomonic projections. *J. Appl. Cryst.* **25**, 294-308.

Case, D.A. & Karplus, M. (1979) Dynamics of ligand binding to heme. *J Mol Biol.* **132**(3), 343-368.

Chu, K., Vojtchovsky, J., McMahon, B.H., Sweet, R.M., Berendzen, J. & Schlichting, I. (2000) Structure of a ligand-binding intermediate in wild-type carbonmonoxy myoglobin. *Nature.* **403**(6772), 921-923.

Cipriani, F., Castagna, J.-C., Wilkinson, C., Oleinek, P. & Lehmann M.S. (1996), Cold Neutron Protein Crystallography using a Large Position- Sensitive Detector based on Image-Plate Technology. *Journal of Neutron Research.* **4** 79

Clifton, I. J., Elder, M. & Hajdu, J. (1991). Experimental strategies in Laue crystallography. *J. Appl. Cryst.* **24**, 267-277.

Cruikshank, D.W.J., Helliwell, J.R. & Moffat, K. (1987). Multiplicity distribution of reflections in Laue diffraction. *Acta Cryst.* **A43**, 656-674.

Cruickshank, D. W. J., Helliwell, J. R. & Moffat, K. (1991). Angular distribution of reflections in Laue diffraction. *Acta Cryst.* **A47**, 352-373.

Cruickshank, D. W. J. (1999). Remarks about protein structure precision. *Acta Cryst.* **D55**, 583-601.

Dauter, Z., Dauter, M., de La Fortelle, E., Bricogne, G., Sheldrick, G.M. (1999). Can anomalous signal of sulfur become a tool for solving protein crystal. *J Mol Biol.* **289**(1), 83-92.

Deacon, A. M., Appleby, T., Bilderback, D. H., Ealick, S. E., Fontes, E. & Thiel, D. J. (1998). Protein crystallography using a multilayer monochromator. *J. Synchrotron Rad.* **5**, 494-496.

Esquerra, R.M., Goldbeck, R.A., Kim-Shapiro, D.B. & Kliger, D.S. (1998). Spectroscopic evidence for nanosecond protein relaxation after photodissociation of myoglobin-CO. *Biochemistry.* **37**(50),17527-17536.

Farber, G. K., Machin, P., Almo, S. C., Petsko, G. A. & Hajdu, J. (1988). X-ray Laue diffraction from crystals of xylose isomerase. *Proc. Nat. Acad. Sc. USA*, **85**(1), 112-115.

Franzen, S., Bohn, B., Poyart, C., DePillis, G., Boxer, S.G. & Martin, J.L. (1995). Functional aspects of ultra-rapid heme doming in hemoglobin, myoglobin, and the myoglobin mutant H93G. *J Biol Chem.* **270**(4), 1718-1720.

French, S. & Wilson, K.S. (1978). On the treatment of negative intensity observations. *Acta Cryst.* **A34**, 517-525.

Fewster, P. F. (1984). Laue orientation and interpretation by microcomputer. *J. Appl. Cryst.* **17**, 265-268.

Glaeser, R., Facciotti, M., Walian, P., Rouhani, S., Holton, J., MacDowell, A., Celestre, R., Cambie, D. & Padmore H. (2000). Characterization of conditions required for X-Ray diffraction experiments with protein microcrystals. *Biophys J*. **78**(6), 3178-3185.

Hajdu, J., Machin, P., Campbell, J. W., Greenhough, T. J., Clifton, I. J., Zurek, S., Gover, S., Johnson, L.N. & Elder, M. (1987a). Millisecond X-ray diffraction and the first electron density map from Laue photographs of a protein crystal. *Nature*, **329**, 178-181.

Hajdu, J., Acharya, K. R., Stuart, D. I., McLaughlin, P. J., Barford, D., Oikonomakos, N. G., Klein, H. & Johnson, L. N. (1987b). Catalysis in the crystal: synchrotron radiation studies with glycogen phosphorylase b. *EMBO J.*, **6**(2), 539-546.

Hajdu, J., Almo, S.C., Farber, G.K., Prater, J.K., Petsko, G.A., Wakatsuki, S., Clifton, I.J. & Fülöp, V. (1991). In: Crystallographic Computing 5: From Chemistry to Biology, edited by D.Moras, A.D. Pojarny & J.C. Thierry, Oxford: Oxford Univ. Press, 29-49.

Hajdu, J. & Andersson, I. (1993), Fast Crystallography and Time-Resolved Structures. *Annu. Rev. Biophys. Biomol. Struct.* **22**, 467-498.

Hao, Q., Campbell, J.W., Harding, M.M. & Helliwell, J.R. (1993). Evaluation of reflection intensities for the components of multiple Laue diffraction spots by direct methods. *Acta Cryst.* **A49**, 528-531.

Hao, Q., Harding M.M. & Campbell, J.W. (1995a). Evaluation of Reflection Intensities for the Components of Multiple Laue Diffraction Spots. III. Using a Real-Space Density Modification Method. *J. Synchrotron Rad.* **2**, 27-30.

Hao, Q., Harding M.M. & Campbell, J.W. (1995b). Determination of dmin and λmin from the Intensity Distributions of Laue Patterns. *J. Appl. Cryst.* **28**, 447-450.

Hartmann, H., Zinser, S., Komninos, P., Schneider, R.T., Nienhaus, G.U. & Parak F. (1996). X-ray structure determination of a metastable state of carbonmonoxy myoglobin after photodissociation. *Proc Natl Acad Sci USA*. **93**(14), 7013-7016.

Hedman, B., Hodgson, K. O., Helliwell, J. R., Liddington, R. & Papiz M. Z. (1985). Protein microcrystal diffraction and the effects of radiation damage with ultra-high-flux synchrotron radiation. *Proc. Nat. Acad. Sc. USA*, **82** (22), 7604-7607.

Hart, H. V. & Rietman, E. A. (1982). Indexing asymmetrical Laue photographs: application to echinoid calcite. *J. Appl. Cryst.* **15**, 126-129.

Helliwell, J.R., Habash, J., Cruickshank, D.W.J., Harding, M., Greenhough, T.J., Campbell, J.W., Clifton, I.J., Elder, M., Machin, P.A., Papiz, M.Z. & Zurek, S. (1989). The recording and analysis of synchrotron X-radiation Laue diffraction photographs. *J.Appl.Cryst.* **22**, 483-497.

Helliwell, J.R. (1984). Synchrotron X-Radiation Protein Crystallography; Instrumentation, Methods and Applications. *Rep. Prog. Phys.* **47**, 1403-1497.

Helliwell, J.R. (1985). Protein Crystallography with Synchrotron Radiation, *J. Molec. Struct.* **130**, 63-91.

Helliwell, J.R. (1991). Macromolecular Crystallography with Synchrotron Radiation, Cambrige: University Press.

Henry. E.R. (1993) Molecular dynamics simulations of heme reorientational motions in myoglobin. *Biophys J.* **64**(3), 869-85.

Ho, J.X., Snell, E.H., Sisk, R.C., Ruble, J.R., Carter, D.C., Owens, S.M. & Gibson W.M. (1998). Stationary crystal diffraction with a monochromatic convergent X-ray source and application for macromolecular crystal data collection. *Acta Crystallogr.* **D54** (Pt 2), 200-214.

Hohlwein, D. & Mason, S.A. (1981). Neutron photographic measurements of protein single-crystal reflections. *J. Appl. Cryst.* **14**, 24-27.

Howell, P. L., Almo, S. C., Parsons, M. R., Hajdu, J. & Petsko, G. A. (1992). Structure determination of turkey egg-white lysozyme using Laue diffraction data. *Acta Cryst.* **B48**, 200-207.

Jones, T.A., Zou, J.Y., Cowtan, S.W & Kjeldgaard, M. (1991), Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Cryst.* **A47**, 110-119.

Kabsch, W. (1988). Automatic indexing of rotation diffraction patterns. *J. Appl. Cryst.* **21**, 67-72.

Kachalova, G. S., Popov, A. N. & Bartunik, H. D. (1999). A steric mechanism for inhibition of CO binding to heme proteins. *Science* **284** (5413), 473-476.

Kachalova, G. S., Popov, A. N. & Bartunik, H. D. (2001). In prep.

Kahn, R., Fourme, R., Gadet, A., Janin, J., Dumas, C. & Andre, D. (1982). Macromolecular crystallography with synchrotron radiation: photographic data collection and polarization correction. *J. Appl. Cryst.* **15**, 330-337.

Kosciesza, D.& Bartunik, H. D. (1999). Extraction of single bunches of synchrotron radiation from storage rings with an X-ray chopper based on a rotating mirror. *J. Synchrotron Rad.* **6**, 947-952.

Kosciesza, D. (2000). PhD thesis. Univ. Hamburg.

Kheiker, D. M. & Nekrasov, Yu. V. (1970). On the Structture Factor Estimates from Intensity Measurements which are Comparable to their Errors. *Instrum. Meth. X-Ray Anal.* (Apparatura i Metody Rentgenovskogo Analysa) 7, 23-28 (in russian).

Laugier, J., Rolland, G. & Emeriau, J. (1989). Dispositif d'orientation rapide de cristaux par la methode de Laue. *J. Appl. Cryst.* **22**, 431-434.

Lebedev, A., Bourenkov, G.P. & Bartunik, H.D. (1998). A General Likelihood Function for MAD/MIR Phasing. DGK 6. Jaherstagung, Karlsruhe.

Lehmann, M. S. & Larsen, F. K. (1974). A method for location of the peaks in step-scan measured Bragg reflexions. *Acta Cryst.* **A30**, 580-584.

Lim, M., Jackson, T.A. & Anfinrud, P.A. (1997). Ultrafast rotation and trapping of carbon monoxide dissociated from myoglobin. *Nat Struct Biol.* **4**(3), 209-214.

Lunin, V. Yu. & Skovoroda, T. P. (1995). R-free likelihood-based estimates of errors for phases calculated from atomic models. *Acta Cryst.* **A51**, 880-887.

Materlik, G. & Tschentscher,Th. (eds.) (2001). TESLA Technical Design Report, Part V, DESY Report TESLA-FEL 2001-05, Hamburg.

Marquart, M., Walter, J., Deisenhofer, J., Bode, W. & Huber, R. (1983). The geometry of the reactive site and of the peptide groups in trypsin, trypsinogen and its complexes with inhibitors. *Acta Cryst.* **B39**, 480.

Moffat, K., Szebenyi, D.M.E. & Bilderback, D.H. (1984). X-ray laue diffraction from protein crystals. *Science* **223**, 1423-1425.

Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). Refinement of Macromolecular Structures by the Maximum-Likelihood Method. *Acta Cryst.* **D53**, 240-255.

Nave, C. (1995). Radiation damage in protein crystallography. *Radiation Physics and Chemistry* **45** (3), 483-490.

Olson, J.S. & Phillips, G.N. (1996). Jr. Kinetic Pathways and Barriers for Ligand Binding to Myoglobin. *J Biol Chem.* **271**(30), 17593-17596.

Ostermann, A., Waschipky, R., Parak, F.G. & Nienhaus, G.U. (2000). Ligand binding and conformational motions in myoglobin. *Nature* **404**(6774), 205-208.

Pannu, N. S. & Read, R. J. (1996). Improved Structure Refinement Through Maximum Likelihood. *Acta Cryst.* **A52**, 659-668.

Perman, B., Srajer, V., Ren, Z., Teng,T., Pradervand, C., Ursby, T., Bourgeois, D., Schotte, F., Wulff, M., Kort, R., Hellingwerf, K. & Moffat K. (1998), Energy transduction on the nanosecond time scale: early structural events in a xanthopsin photocycle. *Science* **279**(5358), 1946-1950.

Ploc, R. A. (1978). Indexing back-reflection Laue patterns by computer. *J. Appl. Cryst.* **11**, 713-715.

Popov, A.N. & Bartunik, H.D., (1996). Unpublished results.

Popov, A.N. (2000). Privat communication.

Ravelli, R. B. G., Hezemans, A. M. F., Krabbendam, H. & Kroon, J. (1996). Towards Automatic Indexing of the Laue Diffraction Pattern. *J. Appl. Cryst.* **29**, 270-278.

Read, R. J. (1986). Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta Cryst.* **A42**, 140-149.

Ren, Z. & Moffat, K. (1995a). Quantitative Analysis of Synchrotron Laue Diffraction Patterns in Macromolecular Crystallography. *J. Appl. Cryst.* **28**, 461-481.

Ren, Z. & Moffat, K. (1995b). Deconvolution of Energy Overlaps in Laue Diffraction. *J. Appl. Cryst.* **28**, 482-483.

Ren, Z., Bourgeois, D., Helliwell, J. R., Moffat, K., Srajer, V. & Stoddard, B. L. (1999). Laue crystallography: coming of age. *J. Synchrotron Rad.* **6**, 891-917.

Reynolds, C. D., Stowell, B., Joshi, K. K., Harding, M. M., Maginn, S. J. & Dodson, G. G. (1988). Preliminary study of a phase transformation in insulin crystals using synchrotron-radiation Laue diffraction. *Acta Cryst.* **B44**, 512-515.

Riquet, J.-P. & Bonnet, R. (1979). Depouillement par ordinateur des cliches de diffraction obtenus par la methode de Laue. *J. Appl. Cryst.* **12**, 39-41.

Rupp, S. (1991). Diploma Thesis, Univ. Hamburg, Germany.

Sagnella. D.E., Straub, J.E., Jackson, T.A., Lim, M. & Anfinrud PA. Vibrational population relaxation of carbon monoxide in the heme pocket of photolyzed carbonmonoxy myoglobin: comparison of time-resolved mid-IR absorbance

experiments and molecular dynamics simulations. *Proc Natl Acad Sci USA.* **96**(25),14324-14329.

Samudzi, C. T. & Rosenberg, J. M. (1992). Detection of Patterson ghost peaks and shape distortions due to sampling in Patterson functions from incomplete data sets. *J. Appl. Cryst.* **25**, 65-68.

Sheldrik, G.W. (1993). SHELXL-93. Univ. Göttingen.

Schlichting, I., Almo, S.C., Rapp, G., Wilson, K., Petratos, K., Lentfer, A., Wittinghofer, A., Kabsch, W., Pai, E.F., Petsko, G.A. & Goody, R.S. (1990). Time-resolved X-ray crystallographic study of the conformational change in Ha-Ras p21 protein on GTP hydrolysis. *Nature*, **345**, 309-315.

Schlichting, I. & Chu,K. (2000). Trapping intermediates in the crystal: ligand binding to myoglobin. *Curr Opin Struct Biol.* **10**(6), 744-52.

Shrive, A.C., Clifton, I.J., Hajdu, J. & Greenhough, T.G. (1990). Laue film integration and deconvolution of spatially overlapping reflections. *J. Appl. Cryst.* **23**, 169-174.

Sivia, D.S. & David, W.I.F. (1994). Bayesian approach to extracting structure-factor amplitudes from powder diffraction data. *Acta Cryst.* **A50**, 703-714.

Singer, P.T., Smalas, A., Carty. R.P., Mangel, W.F. & Sweet, R.M. (1993), The hydrolytic water molecule in trypsin, revealed by time-resolved Laue crystallography. *Science* **259** (5095), 669-673.

Smith Temple, B. (1989) . PhD Thesis, Cornell Univ., USA.

Srajer. V., Teng, T, Ursby, T, Pradervand, C., Ren, Z., Adachi, S., Schildkamp, W., Bourgeois, D., Wulff, M. & Moffat K. (1996). Photolysis of the carbon monoxide complex of myoglobin: nanosecond time-resolved crystallography. *Science* **274** (5293), 1726-1729.

Srinivasan, R. & Chandrasekaran, R. (1966). *Indian J. Pure Appl. Phys.* **4**, 178.

Srinivasan, R. & Ramachandran, G. N. (1965). Probability distribution connected with structure amplitudes of two related crystals. IV. The distribution of the normalized difference. *Acta Cryst.* **19**, 1003-1007.

Stanley, E. (1955) The distribution of the mean intensity of a finite group of reflexions. *Acta Cryst.* **8**, 356.

Stoddard, B.L., Koenigs, P., Porter, N., Petratos, K., Petsko, G.A. & Ringe, D. (1990). Observation of the light-triggered binding of pyrone to chymotrypsin by Laue x-ray crystallography. *Proc. Natl. Acad. Sci. USA* **88**(13), 5503-5507.

Sygusch, J. & Allaire, M. (1988). Sequential radiation damage in protein crystallography. *Acta Cryst.* **A44**, 443-448.

Szebenyi, D.M., Bilderback, D., LeGrand, A., Moffat, K., Schildkamp, W., Temple, B.S. & Teng T.Y. (1992). Quantitive analysis of Laue diffraction patterns recorded with a 120 ps exposure from an X-ray undulator. *J. Appl. Cryst.* **25**, 414-423.

Teng, T.Y., Srajer, V., Moffat, K., Initial trajectory of carbon monoxide after photodissociation from myoglobin at cryogenic temperatures. *Biochemistry.* **36**(40), 2087-21

Yang, X., Ren, Z. & Moffat, K. (1998). Structure Refinement Against Synchrotron Laue Data: Strategies for Data Collection and Reduction. *Acta Cryst.* **D54**, 367-377.

Ursby, T. & Bourgeois, D. (1997). Improved Estimation of Structure-Factor Difference Amplitudes from Poorly Accurate Data. *Acta Cryst.* **A53**, 564-575.

Vellieux, F.M.D., Hajdu, J., Verlinde, C.L.M.J., Groendijk, H., Read, R.J., Greenhough, T.J., Campbell, J.W., Kalk, K.H., Littlechild, J.A., Watson, H.C. & Hol, W. (1993). Structure of glycosomal glyceraldehyde-3-phosphate dehydrogenase from

Trypanosoma brucei determined from Laue data. *Proc. Natl.Acad. Sci. USA*, **90**, 2344-2359.

Viehmann, H. & Bartunik, H.D. (1993). Unpublished results.

Vriend, G. & Rossmann, M. G. (1987). Determination of the orientation of a randomly placed crystal from a single oscillation photograph. *J. Appl. Cryst.* **20**, 338-343.

Wakatsuki, S. (1993). LEAP, Laue Evaluation Analysis Package, for time-resolved protein crystallography. Proc. CCP4 Study Weekend 29-30 January 1993, SERC Daresbury Laboratory, Warrington, England, 71-79.

Wilson, A.J.C. (1949). The probability distribution of X-ray intensities. *Acta Cryst.* **2**, 318.

Wulff M., Schotte F., Naylor G., Bourgeois D., Moffat K. & Mourou G. (1997). Time-resolved structures of macromolecules at the ESRF: Single-pulse Laue diffraction, stroboscopic data collection and femtosecond flash photolysis, *Nuclear Instruments and Methods in Physics Research*, **A398**, 1, 69-84.

List of Publications

The results of this doctoral thesis were included in the following publications:

Bourenkov G. P., Popov, A. N. & Bartunik, H. D. (1996). Bayesian aproach to Laue difraction analysis and its potential for time-resolved protein crystallography. *Acta Cryst.* **A52**, 797-811.

Burenkov, G. P. & Popov A. N. (1994). Method of automaticaly indexing Laue paterns. *Kristallografiya* **39**, 622-627.

The following publications are related to the scope of the thesis but do not constitute a part of the thesis:

Popov, A. N., Andrianova, M. E., Burenkov, G. P & Kheiker, D. M. (1994). Methods of measuring x-ray diffraction intensities for macromolecular single crystals using a KARD-6 multiwire area detector X-ray diffractometer. *Kristallografiya* **39**, 415-421.

Sulyanov, S. N., Burenkov, G. P. & Kheiker, D. M. (1995). The use of an area-detector X-ray diffractometer in the Rietveld method. *Kristallografiya* **40**, 234-238.

Gomis-Ruth, F. X, Maskos, K., Betz,. M., Bergner, A., Huber, R., Suzuki, K., Yoshida, N., Nagase, H., Brew, K., Bourenkov, G.P., Bartunik, H. & Bode, W. (1997). Mechanism of inhibition of the human matrix metalloproteinase stromelysin-1 by TIMP-1. *Nature* **389 (46)**, 77-81.

Than, M.E., Hof, P., Huber R., Bourenkov, G. P, Bartunik, H. D., Buse, G. & Soulimane, T. (1997). *Thermus thermophilus* cytochrome-c552: a new highly thermostable cytochrome-c structure obtained by MAD phasing. *J. Mol. Biol.* **271**, 629-644.

Aleshin, A. E., Zeng, C. B., Bourenkov, G. P., Bartunik, H. D., Fromm, H. J. & Honzatko, R. B. (1998). The mechanism of regulation of hexokinase: new insights from the crystal structure of recombinant human brain hexokinase complexed with glucose and glucose-6-phosphate. *Structure* **6**, 39-50.

Maskos, K., Fernandez-Catalan, C., Huber, R., Bourenkov, G. P., Bartunik, H. D., Ellestad, G. A., Reddy, P., Wolfson, M. F., Rauch, C., T., Castner, B. J., Davis, R., Clarke, H. R. G., Petersen, M., Fitzner, J.N., Cerretti, D. P., March C. J., Paxton, R. J., Black, R. A. & Bode, W. (1998). Crystal structure of the catalytic domain of human tumor necrosis factor-alpha-converting enzyme. *Proc. Nat. Acad. Sc. USA* **95**, 3408-3412.

Dias, J. M., Than, M. E., Humm, A., Huber, R., Bourenkov, G. P., Bartunik, H.D., Bursakov, S., Calvete, J., Caldeira, J., Carneiro, C., Moura, J. J. G., Moura, I. & Romao M. J. (1999) Crystal structure of the first dissimilatory nitrate reductase at 1.9 Å solved by MAD methods. *Struct. fold. des.* **7**, 65-79.

Einsle, O., Messerschmidt, A., Stach, P., Bourenkov, G. P., Bartunik, H. D., Huber, R. & Kroneck, P. M. H. (1999). Structure of cytochrome c nitrite reductase. *Nature* **400**, 476-480.

Grabarse, W., Vaupel, M., Vorholt, J. A., Shima, S., Thauer, R. K., Wittershagen, A., Bourenkov, G., Bartunik, H. D. & Ermler, U. (1999). The crystal structure of methenyltetrahydromethanopterin cyclohydrolase from the hyperthermophilic archaeon *Methanopyrus kandleri. Struct. Fold. Des.* **7**, 1257-1268.

Macedo-Ribeiro, S., Bode, W., Huber, R., Quinn-Allen, M. A., Kim, S.W., Ortel, T.L., Bourenkov, G.P., Bartunik, H.D., Stubbs, M.T., Kane, W.H. & Fuentes-Prior, P. (1999). Crystal structures of the membrane-binding C2 domain of human coagulation factor V. *Nature* **402**, 434-439.

Strobl. S., Fernandez-Catalan, C., Braun, M., Huber, R., Masumoto, H., Nakagawa K., Irie, A., Sorimachi, H., Bourenkow, G., Bartunik, H., Suzuki, K. & Bode, W.. (2000).

The crystal structure of calcium-free human m-calpain suggests an electrostatic switch mechanism for activation by calcium. *Proc. Nat. Acad. Sc. USA* **97**, 588-592.

Wahl, M. C., Bourenkov, G. P., Bartunik, H.D. & Huber, R. (2000). Flexibility, conformational diversity and two dimerization modes in complexes of ribosomal protein L12. *EMBO J.* **19**, 174-186.

Masumoto, J., Nakagawa, K., Irie, S., Sorimachi, H., Suzuki, K., Bourenkov, G. P., Bartunik, H., Fernandez-Catalan, C., Bode, W. & Strobl, S. (2000). Crystallization and preliminary X-ray analysis of recombinant full-length human m-calpain. *Acta Cryst.* **D56**, 73-75.

Bochtler, M., Hartman, C., Song, H. K., Bourenkov, G. P., Bartunik, H. D. & Huber, R. (2000). The structures of HslU and ATP-dependent protease HslU-HslV. *Nature* **403,** 800-805.

Clausen, T., Schlegel, A., Peist, R., Schneider, E., Steegborn, C., Chang, Y. S., Haase, A., Bourenkov, G. P., Bartunik, H. D. & Boos, W. (2000). X-ray structure of MalY from *Escherichia coli*: a pyridoxal 5'-phosphate-dependent enzyme acting as a modulator in mal gene expresion. *EMBO J.* **19**, 831-842.

Aleshin, A. E., Kirby, C., Liu, X. F., Bourenkov, G. P., Bartunik, H. D., Fromm H. J. & Honzatko, R. B. (2000). Crystal structures of mutant monomeric hexokinase I reveal multiple ADP binding sites and conformational changes relevant to alosteric regulation. *J. Mol. Biol.* **296**, 1001-1015.

Burkhard, P., Kammerer, R. A., Steinmetz, M. O., Bourenkov, G. P. & Aebi U. (2000). The coiled-coil triger site of the rod domain of cortexilin I unveils a distinct network of interhelical and intrahelical salt bridges. *Struct. fold. des*. **8**, 223-230.

Kaiser, J. T., Clausen, T., Bourenkow, G. P., Bartunik, H. D., Steinbacher, S. & Huber, R. (2000). Crystal structure of a Nifs-like protein from *Thermotoga maritima*: Implications for iron sulphur cluster asembly. *J. Mol. Biol.* **297**, 451-464.

Scheufler, C., Brinker, A., Bourenkov, G., Pegoraro, S., Moroder, L., Bartunik, H. D., Hartl, F. U. & Moarefi, I. (2000). Structure of TPR domain-peptide complexes : Critical elements in the asembly of the Hsp70-Hsp90 multichaperone machine. *Cell* **101**, 199-210.

Soulimane, T., Buse, G., Bourenkov, G. P., Bartunik, H. D., Huber. R. & Than, M. E. (2000). Structure and mechanism of the aberrant ba(3)-cytochrome c oxidase from *Thermus thermophilus. EMBO J.* **19**, 1766-1776.

Merinov, B., Bourenkov, G. & Bismayer, U. (2000). High-temperature superionic phase of mixed proton conductor [Rb-0.57(NH4)(0.43)](3)H(SeO4)(2): dynamic twinning and anomalous display of dynamicaly disordered hydrogen atoms. *Phys. stat. sol.* **B218**, 365-378.

Muller, J., Lapko, A., Bourenkov, G., Ruckpaul, K. & Heinemann, U. (2001). Adrenodoxin reductase-adrenodoxin complex structure suggests electron transfer path in steroid biosynthesis. *J. Biol. Chem.* **276**, 2786-2789.

Worbs, M., Bourenkov, G.P., Bartunik, H.D., Huber R. & Wahl, M.C. (2001). An extended RNA binding surface through arayed S1 and KH domains in transcription factor NusA. *Mol. Cell* **7**, 1177-1189.

Bosch, J., Tamura, T., Bourenkov, G., Baumeister, W. & Essen, L. O. (2001). Purification, crystalization, and preliminary X-ray difraction analysis of the tricorn protease hexamer from *Thermoplasma Acidophilum*. *J. Struct. Biol.* **134**, 83-87.

Blume, H., Bosecke, P., Bourenkov, G.P., Kosciesza, D. & Bartunik, H.D. (2001). The protein crystalography beamline BW6 at DORIS - automatic operation and high-throughput data collection. *Nucl. Instr. Meth.* **A467**, 1358-1362.

Fritz, G., Roth, A., Schiffer, A., Buchert, T., Bourenkov, G.P., Bartunik, H. D, Huber, H., Stetter, K.O., Kroneck, P.M.H., & Ermler, U. (2002). Structure of adenylylsulfate reductase from the hyperthermophilic A*rchaeoglobus fulgidus* at 1.6-Å resolution. *Proc. Nat. Ncad. Sci. USA* **99**, 1836-1841.

Garrido-Franco, M., Ehlert, S., Messerschmidt, A., Marinkovic, S., Huber, R., Laber, B., Bourenkov, G.P. & Clausen, T. (2002) Structure and function of threonine synthase from yeast. *J. Biol. Chem.* **277**,12396-12405.

Than, M. E., Henrich, S., Huber, R., Ries, A., Mann, K., Timpl, R., Bourenkov, G. P., Bartunik, H. D. & Bode W. (2002). The 1.9-Å crystal structure of the noncolagenous (nc1) domain of human colagen IV shows stabilization via a novel type of covalent met-lys cros-link. *Proc. Nat. Acad. Sc. USA* **99**, 6607-6612.

Karain, W., Bourenkov, G. P. & Bartunik, H. D. (2002). Automated mounting, centering and screening of crystals for high-throughput protein crystallography. *Acta. Cryst.* **D58,** *in press.*