CZECH TECHNICAL UNIVERSITY IN PRAGUE
Faculty of Nuclear Sciences and Physical Engineering
Department of Solid State Engineering

Optimization of computational methods of macromolecular crystallography in studies of biotechnological proteins

DISSERTATION THESIS

Author: **Martin Malý**
Supervisor: **Petr Kolenko**
Supervisor Specialist: **Jan Dohnálek**
Year: **2022**

Bibliografický záznam

Název práce:

Optimalizace výpočetních metod makromolekulární krystalografie při studiu biotechnologických proteinů

Autor:	Martin Malý Katedra inženýrství pevných látek Fakulta jaderná a fyzikálně inženýrská České vysoké učení technické v Praze
Studijní program:	Aplikace přírodních věd
Studijní obor:	Fyzikální inženýrství
Typ práce:	Disertace
Školitel:	Doc. Ing. Petr Kolenko, Ph.D. Katedra inženýrství pevných látek Fakulta jaderná a fyzikálně inženýrská České vysoké učení technické v Praze
Školitel-specialista:	Ing. Jan Dohnálek, Ph.D. Institute of Biotechnology Czech Academy of Sciences
Akademický rok:	2022/2023
Počet stran:	122
Klíčová slova:	makromolekulární krystalografie; rentgenová difrakce; párové upřesňování; FAD-dependentní monooxygenasa; antibiotická rezistence

Bibliographic entry

Title:
Optimization of computational methods of macromolecular crystallography in studies of biotechnological proteins

Author:	Martin Malý Department of Solid State Engineering Faculty of Nuclear Sciences and Physical Engineering Czech Technical University in Prague
Degree Programme:	Application of Natural Sciences
Field of Study:	Physical Engineering
Thesis Type:	Dissertation
Supervisor:	Doc. Ing. Petr Kolenko, Ph.D. Department of Solid State Engineering Faculty of Nuclear Sciences and Physical Engineering Czech Technical University in Prague
Supervisor Specialist:	Ing. Jan Dohnálek, Ph.D. Institute of Biotechnology Czech Academy of Sciences

Academic Year:	2022/2023
Number of Pages:	122
Keywords:	macromolecular crystallography; X-ray diffraction; paired refinement; FAD-dependent monooxygenase; antibiotic resistance

Abstrakt

Makromolekulární krystalografie patří k fundamentálním metodám strukturní biologie, která mimo jiné přispívá k pochopení biochemických procesů. Tato práce se věnuje jak optimalizaci postupů samotné metody, tak její aplikaci při studiu biotechnologických proteinů, které souvisí s mechanismy antibiotické rezistence.

Práce představuje vědecký program *PAIREF*, který automatizuje protokol párového upřesňování krystalových struktur. Výsledky této analýzy určují optimální difrakční limit dat pro vysoké rozlišení s využitím strukturního modelu. Volba tohoto parametru má přitom velký vliv na kvalitu výsledného strukturního modelu. Vylepšení modelů programem *PAIREF* je v této práci ukázáno na řadě referenčních difrakčních dat.

Dále je zde prezentována strukturní analýza enzymu modifikujícího tetracykliny z patogenní bakterie *Stenotrophomonas maltophilia*, která je rezistentní vůči širokému spektru antibiotik. Naše výsledky ukazují, že tento enzym dokáže modifikovat tetracyklinová antibiotika. Práce detailně popisuje postup určení krystalové struktury tohoto enzymu. Celková struktura je velice blízká enzymové rodině tetracyklinových destruktas, avšak kompozice aktivního místa je z významné části odlišná a unikátní.

Abstract

Macromolecular crystallography is one of the fundamental methods of structural biology, which particularly contributes to the understanding of biochemical processes. The thesis focuses on both the optimization of the procedures of the method itself and its application to the study of biotechnological proteins related to the mechanisms of antibiotic resistance.

The thesis presents the scientific program *PAIREF*, which performs automated paired refinement of crystal structures. The results of this analysis determine the optimal high-resolution diffraction limit of data. The choice of this parameter has generally a significant impact on the quality of a structure model. The application of *PAIREF* proves to be beneficial on several diffraction datasets presented in the thesis.

Furthermore, the structural analysis of the tetracycline-modifying enzyme from the pathogenic bacterium *Stenotrophomonas maltophilia*, which is resistant to a wide range of antibiotics, is reported. Our results indicate that this enzyme is capable of the modification of tetracycline antibiotics. The thesis describes in detail the process of determination of the enzyme crystal structure. The overall fold is very similar to the enzyme family of tetracycline destructases, but the composition of the active site is markedly different and unique.

Čestné prohlášení

Prohlašuji, že jsem tuto práci vypracoval samostatně a použil jsem k tomu pouze uvedené zdroje, literaturu a software.

Nemám závažný důvod proti užívání tohoto školního díla ve smyslu § 60 Zákona č.121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon).

Declaration

I declare that I carried out this thesis independently, and only with the cited sources, literature and software.

V Praze, říjen 2022
Prague, October 2022

……...……

Martin Malý

Acknowledgement

I would like to thank many people in their native language:

Především chci poděkovat své rodině a Janče za velkou podporu během celého studia a přípravy této práce.

Děkuji Petrovi Kolenkovi a Honzovi Dohnálkovi za tu spoustu dovedností, které mě naučili, jejich neúnavnou podporu, ať už se zrovna dařilo nebo ne, a důvěru svěřit mi úplně nové výzkumné projekty od jejich samotného počátku.

Ich danke Kay Diederichs für sehr schöne und fruchtbringende Zusammenarbeit.

Děkuji všem svým kolegům z laborky, Jarmile Duškové, Tomášovi Kovaľovi, Tereze Skálové, Márii Trundové, Leoně Švecové, Kristýně Adámkové, Blance Husťákové, Karolíně Špeldové, Jakubovi Hrubému, Michalovi Strnadovi a Jindřichovi Haškovi, že mi byli vždycky ochotní pomoct a poradit.

Děkuji pracovníkům Centra molekulární struktury, BTÚ AV ČR, Honzovi Stránskému, Jirkovi Pavličkovi, Petrovi Pompachovi a Páje Vaňkové za spoustu odvedené práce v souvislosti s mými projekty. Дзякуй таксама Таццяне Чарнавец і Вользе Дзмітрук.

Děkuji Ladislavu Kalvodovi za velkou podporu během celého mého studia na KIPL FJFI ČVUT.

I thank all the staff at beamline 14 at the BESSY II synchrotron (Helmholtz Zentrum Berlin, Berlin, Germany).

I would like to thank for funding: This work was supported by the Ministry of Education, Youth and Sports CR from the ERDF fund – project CAAS (CZ.02.1.01/0.0/0.0/16_019/0000778) to the Faculty of Nuclear Sciences and Physical Engineering, Czech Technical University in Prague, and projects ELIBIO (CZ.02.1.01/0.0/0.0/15_003/0000447), CIISB4HEALTH (CZ.02.1.01/0.0/0.0/ 16_013 and BIOCEV (CZ.1.05/1.1.00/02.0109) to the Institute of Biotechnology AS CR; by the institutional support of the Institute of Biotechnology AS CR (86652036); by the Grant Agency of the Czech Technical University in Prague (SGS19/189/OHK4/3T/14 and SGS22/114/OHK4/2T/14). I also acknowledge CMS-Biocev (Biophysical techniques, Crystallization, Diffraction, Structural mass spectrometry) supported by the Ministry of Education, Youth and Sports CR (LM2015043 and LM2018127).

Content

Introduction

Macromolecular crystallography is a robust experimental method for the determination of atomic structures of biological macromolecules. Together with cryo-electron microscopy and nuclear magnetic resonance, it belongs to the key methods of structural biology. The gained knowledge allows understanding of metabolic processes of cells and is fundamental for various fields in science, medicine and industry. Generally, the crystallographic methods can reveal detailed structural information. In 2022, the number of structures using X-ray crystallography deposited in the Protein Data Bank (Berman, 2000) exceeded 160 000.

However, the method has two serious bottlenecks: crystallization and the phase problem. The latter has been partially overcome with recent developments in computational and prediction software, *e.g. AlphaFold* (Jumper *et al.*, 2021) that can provide suitable model templates. Nevertheless, experimental phasing still plays an important role, especially in projects investigating proteins with a novel fold.

The continual progress in experimental instrumentation, including the 4th generation synchrotrons, free-electron lasers (Hwu and Margaritondo, 2021) and hybrid-photon-counting detectors (Brönnimann and Trüb, 2016), allows solving projects with increasing complexity. Consequently, the methodology and related scientific software have to be continually developed and improved.

In the research conducted during my doctoral studies, I had two major aims of the study. Thus, the thesis consists of a theoretical part and two parts with the presented results:

- **Part 1) Theoretical part:** Principles of macromolecular crystallography and related biophysical methods. Introduction to the FAD-dependent monooxygenases and tetracycline destructases.
- **Part 2) Development of methods:** Optimization and automation of computational methods of macromolecular crystallography, especially the development of a new software providing *paired refinement*.
- **Part 3) Studies of biotechnological proteins:** Structure-function analysis of biotechnologically relevant proteins, particularly the tetracycline-modifying enzyme from *Stenotrophomonas maltophilia* (NCBI Reference Sequence WP_049406473; Arita *et al.*, 2021; O'Leary *et al.*, 2016) including the application of the developed tools during the data processing.

The first part of the thesis covers the theory of X-ray scattering techniques and biophysical methods. The enzymatic family of FAD-dependent monooxygenases is introduced. The second part deals primarily with the development of *PAIREF*, a

software tool that provides a comprehensive analysis of the optimal resolution cutoff for diffraction data during structure model determination. All the results stated in the second part of the thesis were published in peer-reviewed academic journals, except of a short appendix that summarizes recent developments. The *PAIREF* program automates the *paired refinement* protocol that allows the linking of quality of a structure model and measured diffraction data (Diederichs and Karplus, 2013; Karplus and Diederichs, 2012, 2015). The procedure is nowadays generally accepted as the 'golden standard' for the high-resolution diffraction limit estimation (Helliwell, 2022) which is further supported by several arguments included in the thesis.

The work on the program is presented in two attached scientific articles (Malý *et al.*, 2020, 2021). In addition, a description of the feature developed very recently, an automatic decision-making algorithm, is attached in the appendix. *PAIREF* attracted the attention of the scientific community and has been recently included in the *CCP*4 version 8.0 (Krissinel *et al.*, 2018; Winn *et al.*, 2011), a large software package for macromolecular crystallography.

Moreover, I was involved in the development of *SHELIXIR* (Kolenko *et al.*, 2021), a software tool that provides automated experimental phasing with *SHELX C*/*D*/*E* (Sheldrick, 2008, 2010; Usón and Sheldrick, 2018). The corresponding scientific article is included in this thesis.

The third part of the thesis is concerned with the structure-function analysis of a particular FAD-dependent monooxygenase: the tetracycline-modifying enzyme from *Stenotrophomonas maltophilia*, further denoted as *Sm*TetX (NCBI Reference Sequence WP_049406473; Arita *et al.*, 2021; O'Leary *et al.*, 2016). Antibiotic resistance has become a serious global problem (Lewis, 2020; Murray *et al.*, 2022). *S. maltophilia*, a Gram-negative bacterium from the Xanthomonadales order, exhibits broad resistance to most drugs including tetracycline antibiotics (Brooke, 2012; Gajdács and Urbán, 2019; Gil-Gil *et al.*, 2020; Lira *et al.*, 2017; Sánchez, 2015; Xie *et al.*, 2021). Thus, medical treatment of infections caused by this emerging pathogen is very inefficient (Brooke, 2014; Chang *et al.*, 2015; Esposito *et al.*, 2017). A better understanding of this phenomenon includes studies of proteins involved in antibiotic resistance.

S. maltophilia codes for the *Sm*TetX protein possessing close sequence homology with tetracycline destructases, the enzymatic family responsible for the inactivation of tetracycline antibiotics (Fang *et al.*, 2020; Markley and Wencewicz, 2018). This suggests that *Sm*TetX could be involved in antibiotic resistance as well. Our results indicate *Sm*TetX is indeed capable of the inactivation of oxytetracycline. Moreover, the crystal structure reveals novel features in the active site of the enzyme.

Part 1)

Theoretical part

1.1 Scattering of X-ray radiation

For investigation of atomic structures, electromagnetic radiation in the X-ray range, electrons or neutrons are usually used (Helliwell, 2021). Whereas X-rays and electrons interact with charged particles – electrons and nuclei, neutron radiation predominantly with nuclei. These types of radiation also vary in penetration depth which is approximately less than one micrometer for electrons, several micrometers for X-rays but more than a centimeter for neutrons, causing different requirements on sample size (Valvoda *et al.*, 1992). This thesis is particularly focused on the application of X-ray radiation used in macromolecular crystallography and small-angle scattering measurements.

An incident X-ray plane wave transfers its energy to interacting matter that causes vibrations of electrons and nuclei. Thus, the particles became sources of the secondary waves detected during experiments. Inelastic Compton scattering is responsible for background noise signals. However, elastic Thomson scattering, consisting of photons with the same energy as in the incident beam, possesses useful information about the structure of the sample.

The intensity I of elastically scattered radiation on a free charged particle in the distance r is described with the relation

$$I = I_0 \left(\frac{q^2}{4\pi\epsilon_0 m c^2} \right)^2 \frac{1}{r^2} \frac{1+\cos^2(2\theta)}{2}, \tag{1}$$

where I_0 is the intensity of incident radiation, m is a particle weight, q is a particle charge and 2θ is the scattering angle, *i.e.* the angle between the directions of the incident and scattered beams (Kraus, 1985). For X-rays, the contribution of nuclei can be usually neglected as the intensity is inversely proportional to a particle mass, thus, only the contribution of radiation scattered by electrons is considered.

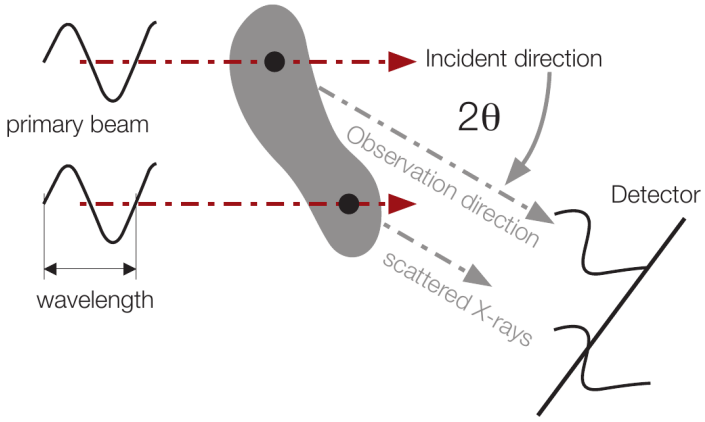


Figure 1: Scattering of a beam of X-ray radiation on matter and the consequent interference that depends on the distance and orientation of the emitting atoms (Schnablegger and Singh, 2017).

When the incident plane wave interacts with multiple atoms, the multiple scattered waves are in superposition (Figure 1). Thus, an interference pattern can be detected (Schmidt, 1995).

The direction of the scattered X-ray waves depends on the radiation wavelength λ, the scattering angle 2θ and the distance r of the atoms from each other. This is usually described with the scattering vector $\mathbf{q} = \mathbf{k_0} - \mathbf{k}$, where $\mathbf{k_0}$ and $\mathbf{k}$ are vectors with a length of $2\pi/\lambda$ in the direction of the incident and scattered beams, respectively, together forming the scattering angle 2θ. The length of the scattering vector, usually stated in the cm^{-1} unit, is

$$|\mathbf{q}| = q = \frac{4\pi}{\lambda}\cdot\sin(2\theta). \tag{2}$$

1.2 Small-angle X-ray scattering (SAXS)

In this method, the distribution of the scattered radiation is measured in order to reveal information about the size, shape and possible aggregation of analyzed particles (Schnablegger and Singh, 2017). For isotropic systems, the distribution usually becomes scattering-angle dependent. The intensity for a system of monodisperse particles in solvent, *i.e.* number of scattered photons per unit time relative to the incident flux σ, per unit solid angle at q and per unit volume Ω, is described with the relation

$$\frac{d\sigma(q)}{d\Omega} = I(q) = I_0\cdot N\cdot(\Delta\rho)^2\cdot V^2\cdot P(q)\cdot S(q), \tag{3}$$

where q is the length of the scattering vector introduced in Equation (2), I_0 is the intensity of incident beam, N is the number of particles per unit volume, $\Delta\rho$ is the contrast, *i.e.* the difference between scattering electron density and solvent electron density, V is the volume of each particle, $P(q)$ denotes the form factor and $S(q)$ the structure factor (Pedersen, 1997).

The form factor $P(q)$ represents the scattering intensity of a single particle normalized to the number of excess electrons of one particle. At $q = 0$, the form factor is unity: $P(0) = 1$. The structure factor $S(q)$ takes into account the particle positions relative to each other. Generally, $S(q\rightarrow\infty) = 1$; for dilute systems it can be assumed $S(q) = 1$.

The measured scattering pattern $I(q)$ can be divided into three parts (Figure 2) according to ranges of the length of the scattering vector q: Guinier q range, intermediate q range and Porod q range (Porod, 1951).

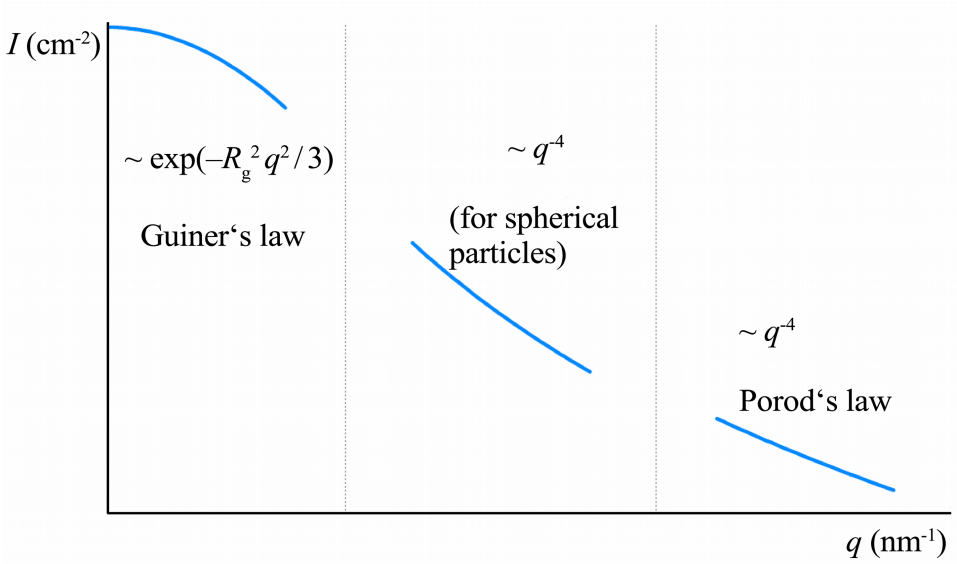


Figure 2: Typical dependency of the scattered intensity I on the length of the scattering vector q, divided into the three characteristic q ranges (Li, 2005).

1.2.1 Guinier q range

At small angles, *i.e.* small q, the function in the Equation (3) can be approximated with a Gaussian curve that gives Guinier's law (Guinier, 1939):

$$I(q) \approx I(0)\cdot\exp\left(\frac{-R_g^2\cdot q^2}{3}\right), \tag{4}$$

where $I(0)$ is the extrapolated zero-angle intensity. R_g denotes the radius of gyration, defined as the root mean square average of the distance from the center-of-mass. Guinier's law is usually represented in logarithmic scale versus q^2 in Guinier plot that allows determination of the parameters R_g and $I(0)$:

$$\ln[I(q)] = \ln[I(0)] - \frac{R_g^2\cdot q^2}{3}. \tag{5}$$

Hence, measured intensities are fit with a linear curve with a slope of $-R_g^2/3$ intersecting the y-axis at a value of $\ln[I(0)]$ according to the relation. Monodisperse samples without aggregation exhibit linearity of the plot near $q = 0$, as is shown in the particular study in Figure 13c.

For a sample with a known concentration c, the molecular weight of the particles M_W can be estimated using the formula derived from Equation (3) for $q = 0$ and a dilute system:

$$M_W = \frac{N_A\cdot I(0)}{c\cdot\Delta\rho_M^2}, \tag{6}$$

where $N_A = 6.022\cdot 10^{23}$ is Avogadro's number, c is the concentration of the particles in weight per unit volume and $\Delta\rho_M$ is the contrast per mass (Feigin and Svergun,

1987; Orthaber *et al.*, 2000). Furthermore, molecular mass can be estimated using several other principles (Mylonas and Svergun, 2007).

1.2.2 Intermediate q range

The decay of intensity in this range indicates an overall shape of the analyzed particles. A decay of q^{-4} is characteristic for particles possessing a spherical shape (*e.g.* globular proteins), whereas of q^{-1} and q^{-2} for a cylindrical and a flat shape, respectively.

As the form factor usually oscillates in this range, it is convenient to calculate Fourier transformation to get the pair-distance distribution function in real space $p(r)$ (Glatter, 1977; Liu and Zwart, 2012):

$$p(r) = \frac{r}{2\pi^2} \int_{q_{min}}^{q_{max}} q\, I(q) \sin(qr)\, \mathrm{d}q\,. \tag{7}$$

This function represents the histogram of distances possibly found within the particle. For globular particles, it exhibits a bell-shaped symmetric peak. Application of the pair-distance distribution function is crucial for the modeling the structure of analyzed particles. An initial structure model is randomly or systematically modified to achieve the best agreement between the theoretically calculated and experimentally determined $I(q)$ and $p(r)$ profiles.

1.2.3 Porod q range

In this range, the intensity decays with $K_P \cdot q^{-4}$ which is called Porod's law (Porod, 1951). The constant K_P contains several instrumental factors. Moreover, it is proportional to the ratio of surface and sample volume, thus, it can be used to estimate the surface area of domains.

The second moment of a scattering profile Q is invariant, *i.e.* it is a universal constant for any SAXS scattering pattern:

$$Q = \int_{-\infty}^{\infty} q^2\, I(q)\, \mathrm{d}q\,. \tag{8}$$

The invariant Q, which is a theoretically well-defined constant, contains the same instrumental factors as the constant K_P. Hence, Q is used for the determination of unknown instrumental factors (Schnablegger and Singh, 2017).

1.2.4 Kratky plot

Kratky plot is often represented in its dimensionless variant: the dependency of $(q \cdot R_g)^2 \cdot I(q)/I(0)$ on $q \cdot R_g$ (Glatter and Kratky, 1982). This allows investigating the flexibility or possible degree of unfolding. Globular well-folded proteins exhibit a bell-shaped peak. Kratky plot is shown in the particular study in Figure 13d.

1.2.5 Measurement and processing of SAXS data

Firstly, datasets of an empty sample holder, of the sample holder filled with deionized water and of the sample holder filled with a solvent, *i.e.* a sample buffer, are measured. Then, samples with a known concentration are measured in the sample holder, followed by a repeated control measurement of the solvent. All the experimental equipment (an X-ray source, optics, sample holder and detector) are in a stable immobile position during the data collection (Svergun *et al.*, 2013).

Every dataset usually consists of tens of two-dimensional scattering patterns. Their processing starts with masking of the pixels possessing shadow from a beamstop or other parts. The individual two-dimensional patterns are integrated into the reduced one-dimensional intensity profiles as a function of the length of the scattering vector. Then, the average of the profiles within the dataset is calculated. In order to get information about the analyzed particles in the solvent, the intensity profile of the solvent is subtracted from the intensity profile of the sample (Konarev *et al.*, 2003).

Finally, the intensity can be normalized on the absolute scale by the flux intensity of the incident beam and the illuminated volume of the sample. For this normalization to cm^{-2}, the intensity profiles of the empty sample holder and the sample holder filled with water are taken into account (Schnablegger and Singh, 2017).

1.3 Diffraction on crystal and structure determination

When a monochromatic plane X-ray electromagnetic wave with the wave vector $\mathbf{k_0}$ interacts with atoms that are placed periodically in a crystal lattice defined with the vectors **a**, **b**, **c**, the elastically scattered waves interfere constructively possessing the wave vector **k**, $|\mathbf{k}| = |\mathbf{k_0}|$, and the scattering angle 2θ, if they fulfill the Laue diffraction conditions:

$$\begin{aligned} \mathbf{a}\cdot(\mathbf{k}-\mathbf{k_0}) &= 2\pi h \\ \mathbf{b}\cdot(\mathbf{k}-\mathbf{k_0}) &= 2\pi k \\ \mathbf{c}\cdot(\mathbf{k}-\mathbf{k_0}) &= 2\pi l\,, \end{aligned} \tag{9}$$

where the reflection indices h, k, l are integers (Valvoda *et al.*, 1992).

Let **H**(h, k, l) be a reciprocal lattice vector defined as

$$\mathbf{H}(h,k,l) = h\cdot\mathbf{a}^* + k\cdot\mathbf{b}^* + l\cdot\mathbf{c}^*. \tag{10}$$

The reciprocal lattice vectors **a***, **b***, **c*** are characterized with the cyclical equations $\mathbf{a}\cdot\mathbf{a}^* = 1$, $\mathbf{a}\cdot\mathbf{b}^* = 0$, $\mathbf{a}\cdot\mathbf{c}^* = 0$ (Helliwell, 2004). Hence, this situation can be described with the scattering vector $\mathbf{q} = \mathbf{k}_0 - \mathbf{k}$ or simpler with its alternate version defined as

$$\mathbf{s} = \frac{\mathbf{k}-\mathbf{k}_0}{2\pi}, \tag{11}$$

which gives (Rossmann, 2001)

$$\begin{aligned} \mathbf{a}\cdot\mathbf{s} &= h \\ \mathbf{b}\cdot\mathbf{s} &= k \\ \mathbf{c}\cdot\mathbf{s} &= l. \end{aligned} \tag{12}$$

Based on these relations, the diffraction conditions are fulfilled when the scattering vector **s** is identical to a vector of reciprocal lattice **H**(h, k, l). Hence, a diffraction pattern is a projection of a reciprocal lattice and the reflection indices h, k, l represent coordinates in the reciprocal space (Kraus, 1985).

The length of the diffraction vector **s** can be expressed as

$$|\mathbf{s}| = \frac{|\mathbf{k}-\mathbf{k}_0|}{2\pi} = \frac{|\mathbf{k}||\mathbf{n}-\mathbf{n}_0|}{2\pi} = \frac{2\pi}{\lambda}\frac{|\mathbf{n}-\mathbf{n}_0|}{2\pi} = \frac{2\sin\theta}{\lambda}, \tag{13}$$

where $\mathbf{n}_0$ and **n** are unit vectors in the directions of vectors $\mathbf{k}_0$ and **k**. The diffraction conditions are geometrically described with the Ewald sphere construction (Figure 3).

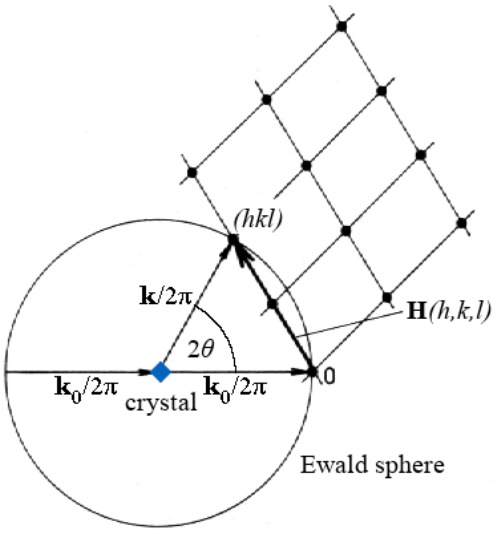


Figure 3: Ewald sphere construction: A crystal is placed in the center of Ewald sphere with a radius of $1/\lambda$. The origin of a reciprocal lattice is located at the point 0 where the primary beam leaves the sphere. While a reciprocal-lattice point (*hkl*) crosses the Ewald sphere, diffraction conditions are met for the crystal planes (*hkl*) (Valvoda *et al.*, 1992).

Thus, the identity $|\mathbf{s}| = |\mathbf{H}(h, k, l)|$ implies Bragg's law describing the diffraction condition

$$2 d_{hkl} \sin\theta = \lambda, \tag{14}$$

where $d_{\mathrm{hkl}} = 1/|\mathbf{H}(h, k, l)|$ is the distance between the crystal planes (hkl) (McPherson, 2009).

1.3.1 Structure factor

The ability of an atom to scatter radiation is described with the atomic scattering factor $f(\mathbf{s})$:

$$f(\mathbf{s}) = \int \rho(\mathbf{r}) \exp(2\pi \mathrm{i}\, \mathbf{r}\cdot\mathbf{s}) \mathrm{d}\mathbf{r}, \tag{15}$$

where $\rho(\mathbf{r})$ is the electron density depending on the Cartesian coordinates $\mathbf{r} = (x, y, z)$ (Valvoda *et al.*, 1992).

Let the atoms be periodically placed in the crystal lattice with the unit cell consisting of N atoms and the diffraction conditions are fulfilled for the crystal planes (hkl). Then the structure factor can describe the atom positions in the unit cell:

$$F_{hkl} = \sum_{j=1}^{N} f_{\mathrm{j}} \exp[2\pi \mathrm{i}(hx_{\mathrm{j}} + ky_{\mathrm{j}} + lz_{\mathrm{j}})], \tag{16}$$

where x_j, y_j and z_j are the coordinates of the j-th atom in the unit cell. The formula implies that the structure factor is a complex number $F_{hkl} = |F_{hkl}| \cdot \exp(2\pi \mathrm{i} \varphi_{hkl})$. In diffraction experiments, the intensities I_{hkl} are measured that correspond to the squared absolute values of the structure factor $I_{hkl} \sim |F_{hkl}|^2$ (Rupp, 2009).

When the structure factor F_{hkl} is known, the electron density that characterizes the structure of an investigated molecule can be calculated as the inverse Fourier transform of F_{hkl} (Drenth and Mesters, 2007):

$$\rho(x, y, z) = \sum_{hkl} \frac{1}{V} |F_{hkl}| \exp[-2\pi \mathrm{i}(xh + yk + zl) + \mathrm{i}\phi_{hkl}], \tag{17}$$

where $\mathbf{r} = (x, y, z)$ are the coordinates in the unit cell and V is the volume of the unit cell. However, the phase information φ_{hkl} is not usually available from the diffraction experiment. Nevertheless, there are several experimental and computational methods for the estimation of φ_{hkl}. The methods of experimental phasing represent single/multiple isomorphous replacement (Weeks *et al.*, 2003) and single/multiple anomalous dispersion, which is based on the monitoring of differences between $|F_{hkl}|$ and $|F_{-h,-k,-l}|$ (Hendrickson *et al.*, 1985). The preparation of high-quality crystals that meets requirements for these experiments is often demanding. Thus, the molecular replacement method is used more often. It will be applied also in this thesis and further explained below.

1.3.2 Diffraction experiment

Diffraction data collection is performed in a single-crystal diffractometer (Figure 4) which mainly consists of an X-ray radiation source, a goniometer and a detector (the X-ray radiation sources and detectors are introduced in a forthcoming section below). The sample – crystal of a macromolecule – is placed in a loop that is magnetically mounted on the goniometer (Drenth and Mesters, 2007). The crystal is usually cooled using a cryostat with nitrogen vapors to a temperature of 100 K which significantly reduces radiation damage (Garman and Schneider, 1997). A primary-beam-stop is placed between the crystal and the detector.

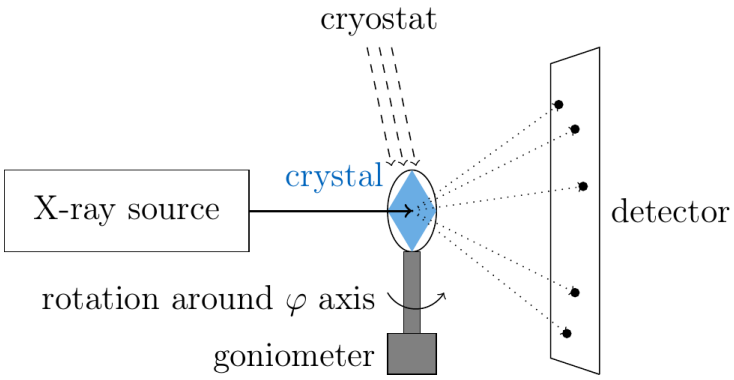


Figure 4: Schematic diagram of diffraction experiment on single-crystal using a φ-axis goniometer (Malý, 2018).

Firstly, a few diffraction images at different crystal rotations are measured to characterize the quality of diffraction of a crystal and to index the observed reflections. If the crystal gives a sufficient diffraction signal, a data collection strategy can be calculated, *e.g.* with the program *iMOSFLM* (Battye *et al.*, 2011). The data collection is subsequently performed with the rotation method: the sample is evenly rotated around the φ axis whereas the radiation source and the detector are in a stable position. Mostly hybrid-pixel detectors are used which allows the reading of the measured values of intensity continuously during a rotation of oscillation angle $\Delta\varphi$, usually several tenth of a degree – denoted as „fine φ-slicing“ (Mueller *et al.*, 2012).

1.3.3 Diffraction data processing

The first step of the process is indexing, *i.e.* assignment of the *hkl* indices to the strong reflections in diffraction patterns. Based on the positions of reflections, unit cell parameters and Laue symmetry group are assigned. The setting of the correct geometry parameters, mainly the crystal-detector distance and the position of primary beam, is crucial for the successful indexing (Marek and Trávníček, 2002).

Then, possible positions of all reflections can be predicted and their intensities are integrated. The algorithm usually performs the profile analysis: shapes of intense reflections are fitted and the obtained typical reflection profiles are applied for

integration of weak reflections. Simultaneously, the unit cell and geometry parameters are refined (Kabsch, 2010a).

The last step is the scaling of the reflection intensities with the crystal symmetry taken into account. Finally, individually observed intensities corresponding to the same reflection *hkl* are merged (Winter *et al.*, 2018).

The quality of diffraction data can be quantified with several indicators, *e.g.* the average intensity-to-noise ratio $I/\sigma(I)$, R_{meas} (Diederichs and Karplus, 1997; Weiss, 2001) or the Pearson correlation coefficient between two randomly chosen subsets of reflections $CC_{1/2}$ (Assmann *et al.*, 2016; Karplus and Diederichs, 2015). While the intensity of reflections decreases with increasing resolution (decreasing d_{hkl}), diffraction data are usually cut at a certain high-resolution level according to criteria based on these statistics to discard reflections containing no useful signal. Most frequently, the following criteria for the reflections in the highest resolution shell are applied: $I/\sigma(I) > 2$ or $CC_{1/2} > 0.50$. However, this conservative approach often leads to the rejection of reflections that still contain a useful signal (Diederichs and Karplus, 2013; Karplus and Diederichs, 2015). Thus, the choice of the high-resolution cutoff should be later checked during the refinement of a structure model performing the paired refinement protocol (Karplus and Diederichs, 2012; Malý *et al.*, 2020).

For diffraction data processing, several programs have been developed; the most used are nowadays *XDS* (Kabsch, 2010b), *DIALS* (Winter *et al.*, 2018) and *iMOSFLM* (Battye *et al.*, 2011). There are also several software tools focused on the intensity scaling and merging: *XSCALE* (Kabsch, 2010b), *AIMLESS* (Evans and Murshudov, 2013) and *STARANISO* (Tickle *et al.*, 2018).

1.3.4 Patterson function

Generally, the Patterson function is fundamental for the methods of phase problem solution. This function is defined as an inverse Fourier transform of the absolute value of structure factor squared. Hence, the phase information is not necessary for the calculation (Drenth and Mesters, 2007):

$$P(\mathbf{u}) = \frac{1}{V}\sum_{hkl} |F_{hkl}|^2 \exp[-2\pi\mathrm{i}(uh+vk+wl)], \tag{18}$$

where $\mathbf{u} = (u, v, w)$ is the vector of coordinates in the unit cell. When we put F_{hkl} from (16) to this formula (18), we get

$$P(\mathbf{u}) = \frac{1}{V}\sum_{hkl}\sum_{i=1}^{N}\sum_{j=1}^{N} f_i f_j \exp\{2\pi\mathrm{i}[h(\Delta x_{ij}-u)+k(\Delta y_{ij}-v)+l(\Delta z_{ij}-w)]\}, \tag{19}$$

where $\Delta\mathbf{r}_{ij} = (\Delta x_{ij}, \Delta y_{ij}, \Delta z_{ij}) = (x_i-x_j, y_i-y_j, z_i-z_j)$. Hence, the Patterson function has an extreme while $\mathbf{u}$ equals $\Delta\mathbf{r}_{ij}$, *i.e.* the positions of the maxima (Patterson vectors) correspond with the typical interatomic distances in the unit cell.

Moreover, the Patterson function is the convolution of the electron density and its inverse (McPherson, 2009). Assuming the unit cell consists of N atoms, it can be expressed with the formula (Marek and Trávníček, 2002)

$$P(\mathbf{u}) = \rho(\mathbf{r}) * \rho(-\mathbf{r}) = \sum_{j=1}^{N} \rho(\mathbf{r}_j)\rho(\mathbf{r}_j + \mathbf{u}), \tag{20}$$

where $\mathbf{r}_j$ is the position of the j-th atom in the unit cell.

Thus, the Patterson function exhibits the same periodicity as the electron density and the maxima are located in positions of interatomic vectors.

1.3.5 Solution of phase problem using molecular replacement

Molecular replacement (Rossmann and Blow, 1962; Rossmann, 1972) is the most rapid method for solving the phase problem. However, a determined structure of a homologous protein is required to be used, thus it is not always possible to apply this method (Drenth and Mesters, 2007).

The principle of the method is to obtain unknown phase information using a similar solved structure. The problem is to place the model molecule in the proper orientation and position in the target unit cell. Thus the process can be divided into two steps: rotation and translation:

$$\mathbf{X}' = \mathbb{C}\mathbf{X} + \mathbf{t}, \tag{21}$$

where $\mathbf{X}$ and $\mathbf{X}'$ are the Cartesian coordinates of the model and the final molecule, respectively, $\mathbb{C}$ is the rotational matrix and $\mathbf{t}$ is the translation vector.

Pairs of atoms belonging to the same molecule generate short self-Patterson vectors whose end-points are located close to the origin (Drenth and Mesters, 2007). The same molecules in different crystal structures have the inner region of the Patterson map equal apart from rotation. Hence, the self-Patterson vectors can be used for finding the proper orientation of the model structure to the unknown molecule as

The rotation function is defined as

$$R(\mathbb{C}) = \int_U P_O(\mathbf{u}) \cdot P_M(\mathbb{C}\mathbf{u})\, \mathrm{d}\mathbf{u}, \tag{22}$$

where U is the volume of Patterson map, where the self-Patterson vectors are located, $P_O(\mathbf{u})$ is the Patterson function based on the measured observed intensities and $P_M(\mathbb{C}\mathbf{u})$ is the rotated Patterson function belonging to the model structure (Evans and McCoy, 2008). Thus, the rotation function has a maximum value for the proper overlap.

Subsequently after the estimation of the rotational matrix $\mathbb{C}$, the translation vector $\mathbf{t}$ is determined using the intermolecular cross-Patterson vectors that are relatively longer than the self-Patterson vectors. Similarly to the previous case, the Patterson

functions of the investigated and the model structure are compared and the maximal overlap depending on the translation vector $\mathbf{t}$ is searched. The level of overlap is expressed by the translation function

$$T(\mathbf{t}) = \int_V P_O(\mathbf{u}) \cdot P_{M;a,b}(\mathbf{u},\mathbf{t})\,\mathrm{d}\mathbf{u}, \tag{23}$$

where $P_{M;a,b}$ are the cross-Paterson vectors of the model structure between the a-th and the b-th molecule in the unit cell with a volume of V (Marek and Trávníček, 2002). Hence, the absolute position of the investigated molecule in the unit cell is then determined.

The agreement between the structure models and the experimental observed data can be quantified with several indicators:

- crystallographic R-value (Stout and Jensen, 1989):

$$R = \frac{\sum_{hkl} \left\| F_{O;hkl} \right| - k \left| F_{C;hkl} \right\|}{\sum_{hkl} \left| F_{O;hkl} \right|}, \tag{24}$$

  where $F_{O;hkl}$ and $F_{C;hkl}$ is the structure factor based on the observed experimental data and structure model, respectively, and k is a scaling factor;

- Pearson correlation coefficient (Vagin and Teplyakov, 1997):

$$\mathrm{CC} = \frac{\sum_{hkl} \langle \left| F_{O;hkl} \right| \left| F_{C;hkl} \right| - \langle \left| F_{O;hkl} \right| \rangle \langle \left| F_{C;hkl} \right| \rangle \rangle}{\sum_{hkl} \sqrt{\langle \left| F_{O;hkl} \right|^2 - \langle \left| F_{O;hkl} \right| \rangle^2 \rangle} \sqrt{\langle \left| F_{C;hkl} \right|^2 - \langle \left| F_{C;hkl} \right| \rangle^2 \rangle}}, \tag{25}$$

  where the angle brackets denote the mean value;

- log likelihood gain (LLG) (Murshudov *et al.*, 1997; Oeffner *et al.*, 2018);
- translation-function Z-score (McCoy *et al.*, 2007).

There are several publicly available programs to perform molecular replacement, *e.g.* *MOLREP* (Vagin and Teplyakov, 1997), *Phaser* (McCoy *et al.*, 2007), and pipelines *MoRDA* (Vagin and Lebedev, 2015) and *MrBUMP* (Keegan and Winn, 2007).

1.3.6 Refinement of structure model

After the phase problem solution, it is possible to calculate the electron density using Equation (17). However, the agreement between the structure factors based on the initial structure model and the observed structure factors is rather poor, especially the phase information is very inaccurate. Thus, refinement of the model structure is necessary (Drenth and Mesters, 2007).

Refinement in macromolecular crystallography is an iterative process including several steps performed in the reciprocal and the real space. The observed absolute values of structure factor for individual independent reflections $|F_{O;hkl}|$ define the computational problem to calculate the amplitudes $|F_{C;hkl}|$ and phases φ_{hkl}. The refined free parameters in the direct space are the following:

- coordinates of individual atoms in the unit cell (three parameters per atom);
- atomic displacement parameter (ADP, sometimes also denoted as a *B*-factor or Debye-Waller factor (Kraus, 1985)); one or six parameters per atom for isotropic or anisotropic refinement, respectively.

Theoretically, the reciprocal space refinement can be carried out with the least squares method that is implemented *e.g.* in program *SHELXL* (Sheldrick, 2015) from the *SHELX* software package (Sheldrick, 2008). However, diffraction data collected up to a very high resolution (ca. 1 Å) are required for the convergence in order to have a sufficiently high ratio between the number of observations and the number of refined parameters. This assumption is often not fulfilled. Thus, the maximum likelihood method, which has a higher radius of convergence, is usually used (Drenth and Mesters, 2007). This approach is implemented in programs *REFMAC*5 (Kovalevskiy *et al.*, 2018; Murshudov *et al.*, 2011), *Phenix.refine* (Adams *et al.*, 2010; Afonine *et al.*, 2012; Liebschner *et al.*, 2019), *BUSTER* (Blanc *et al.*, 2004) and others.

Stereochemical data of the ideal geometry of amino acids, nucleic acids and ligands from the Cambridge Structural Database (Groom *et al.*, 2016) are also taken into account during the refinement. The root mean square deviations of bond lengths and bond angles from their ideal values are monitored.

The agreement between diffraction data and a refined structure model is monitored with the *R*-value, defined in relation (23), which is cross-validated in order to not be biased (Brünger, 1992, 1997): usually 5 % of the observed reflections are randomly selected to form a test set of **free** reflections excluded from the refinement. Thus, the refinement is performed using the remaining 95 % reflections, assigned in a **working** set. These sets of reflections are not correlated. Hence, *R*-values related to these two individual sets of reflections can be calculated: R_{free} and R_{work}. A decrease in R_{free} after the refinement represents an improvement of the model structure.

The real space refinement, *i.e.* manual modifications of the structure model, is also carried out during the refinement process. For this purpose, several programs have been developed, *e.g. Coot* (Emsley *et al.*, 2010) or *UCSF ChimeraX* (Pettersen *et al.*, 2021). An *R*-value of the final refined structure model should be in a range of 10-25 % after the refinement process (depending partially on a diffraction limit).

1.3.7 Paired refinement to determine optimal resolution cutoff

Diffraction data are usually cut at high-resolution using conservative criteria on data quality indicators, *e.g.* $I/\sigma(I)$, $CC_{1/2}$. However, these statistics do not take into account the quality of structure model. Thus, the impact of the previously discarded reflections on the structure model should be checked during the refinement process as they often still contain a useful signal (Karplus and Diederichs, 2015). For this purpose, the paired refinement protocol is carried out (Karplus and Diederichs, 2012). The principle is a stepwise addition of the high-resolution shells into the refinement process with monitoring of statistics related to the structure model, especially the *R*-values. The application of this routine will be explained in the example:

1. Let us assume that diffraction data were processed up to 2.0 Å and then cut at 2.3 Å according to the conservative criterion of $CC_{1/2} > 0.50$ for the highest resolution shell. Thus, the corresponding structure model was refined at 2.3 Å.
2. The next high-resolution shell (2.3-2.2 Å) is added to the refinement process: The structure model is refined at 2.2 Å.
3. The statistics of the pair of structure models, refined using the different cutoffs (2.3 Å and 2.2 Å), are compared. A decrease in R_{free}, or a constant R_{free} with an increase in R_{work} of the model refined at higher resolution denotes model improvement. The *R*-values must be calculated against the same data range to be comparable (*i.e.* at the poorer resolution limit 2.3 Å).
4. Then, the points 2. and 3. are repeated with the next high-resolution shells. The shell 2.2-2.1 Å is added and *R*-values for the corresponding pair of models, refined at 2.2 Å and 2.1 Å, are compared. The same will be then carried out with the final shell 2.1-2.0 Å.

Limited implementation of this routine is included in the programs *PDB-REDO* (Joosten *et al.*, 2014) and *Phenix.refine* (Afonine *et al.*, 2012; Liebschner *et al.*, 2019). In this thesis, we report our software tool *PAIREF* which provides a comprehensive analysis for the determination of the optimal high-resolution limit based on the paired refinement protocol (Malý *et al.*, 2020, 2021).

1.3.8 Structure validation

The final step of the determination of a crystal structure is its validation and quality control. The following issues are particularly monitored:

- Comparison of observed bond lengths and angles with the ideal values from stereochemical libraries (Groom *et al.*, 2016), *e.g.* Ramachandran plot for the monitoring of allowed regions of backbone torsion angles (Ramachandran *et al.*, 1963).

- Interatomic distances to avoid atomic clashes.
- Network of hydrogen bonds, *e.g.* orientation of side chains of the asparagine, glutamine and histidine residues.
- Analysis of rotamers – most frequently observed geometries of side chains.
- Extraordinary low or high values of atomic displacement parameter.
- Agreement between the structure model and experimental data.

The validation can be carried out in *e.g. Coot* (Emsley *et al.*, 2010), *MolProbity* (Williams *et al.*, 2018), *PROCHECK* (Laskowski *et al.*, 1993) and validation service of the Protein Data Bank (Gore *et al.*, 2017; Smart *et al.*, 2018). The final structure that meets the criteria can be deposited to the Protein Data Bank (Berman, 2000; Berman *et al.*, 2003).

1.4 Instrumentation for X-ray scattering experiments

1.4.1 X-ray sources

X-ray radiation is electromagnetic radiation in a range of wavelengths 10^{-9}–10^{-12} m. Generally, the X-ray radiation sources are classified as conventional and synchrotron. The main parameter describing the quality of a beam for physical experiments is the brilliance, defined as the number of photons per second in the beam per millimeter squared of the source surface per milliradian squared in the wavelength spectral range $\Delta\lambda/\lambda = 10^{-3}$ (Mobilio *et al.*, 2015; Wiedemann, 2007).

In the case of the **conventional sources**, *e.g.* X-ray tubes, the radiation is emitted by stopping high-energy electrons in a metal target (Behling, 2015). An X-ray tube consists of a cathode and a positively charged anode at high voltage (ca. tens of kV) in a vacuum. The cathode is heated up by an electric current that leads to thermoemission of electrons. They are accelerated in the electric potential and collide with the anode. Some incident electrons knock an electron from the inner electron shell of an anode atom, whose place is subsequently occupied by an electron from a higher energy shell, and the energy excess is emitted as X-rays. These collisions account for the characteristic component of the emitted radiation spectrum which depends on the anode material (*e.g.* copper, cobalt). Moreover, some incident electrons lose their energy gradually through several collisions that account for the continuous component of the spectrum. Cooling of the anode represents a challenging technical problem and can be improved with the use of a rotating anode or metal-jet-liquid anode (Espes *et al.*, 2016; Otendal *et al.*, 2008).

A **synchrotron storage ring** is a circular particle accelerator. Electrons are firstly accelerated to almost the speed of light and injected into the storage ring (Koch *et al.*, 1983). The circular trajectory is defined by bending magnets. While the electrons change the direction of movement, electromagnetic radiation is emitted at the tangent direction to the storage ring and utilized by beamlines. Parameters of the beam are defined with X-ray mirrors, monochromators, filters and collimators. Apart from the bending magnets, the storage ring contains also insertion devices, wigglers and undulators, that are composed of periodically placed magnets and provide very intense radiation (Hwu and Margaritondo, 2021). The wavelength of the emitted synchrotron radiation can be precisely tuned which is required *e.g.* for the experimental phasing experiments. The beam possesses many orders of magnitude higher brilliance (over 10^{18} photons $\cdot$ s^{-1} $\cdot$ mm^{-2} $\cdot$ mrad^{-2} / 0.1% bandwidth) in comparison with the conventional sources. Furthermore, it is notable for the high level of polarization and collimation.

Recently, **free-electron laser** facilities have been developed, basing in undulator. Femto-second electron bunches go through linear accelerators, long in a range of kilometers, and emit electromagnetic radiation in a special robust undulator or wiggler. The beam is very brilliant and intense and its pulse character especially allows for performing time-resolved experiments (Margaritondo and Rebernik Ribic, 2011).

1.4.2 X-ray detectors

There are several methods and corresponding devices for the X-rays detection, *e.g.* Geiger-Müller counters, ionization chambers, proportional detectors, photographic film, image plates, semiconductor detectors *etc*. In diffraction experiments of macromolecular crystallography, two-dimensional hybrid-pixel counting detectors are used most frequently nowadays at synchrotrons, *e.g.* Dectris PILATUS (Kraft *et al.*, 2009; Loeliger *et al.*, 2012), Decris EIGER (Casanas *et al.*, 2016; Dinapoli *et al.*, 2011; Johnson *et al.*, 2014) and JUNGFRAU (Leonarski *et al.*, 2020). Every individual pixel represents an independent semiconductor sensor that detects X-rays directly owing to the photoelectric effect. The electric pulse is subsequently processed by coupled integrated circuits. The detectors operate in single-photon-counting mode. Detection of noise is suppressed by the application of photon energy thresholds (Brönnimann and Trüb, 2016).

1.5 Selected biophysical methods for investigation of biological macromolecules

1.5.1 Characterization of protein sample

After several steps of purification, the obtained protein sample is characterized using biophysical techniques in order to choose an optimal storage buffer and inspect proper protein folding and stability (Kwan *et al.*, 2019).

Nanoscale differential scanning fluorimetry determines the thermal stability and the melting temperature of proteins (Magnusson *et al.*, 2019; Pantoliano *et al.*, 2001). Changes in intrinsic protein fluorescence are measured as a function of temperature as protein unfolds. The monitored wavelengths (330 nm and 350 nm) correspond to the fluorescence of tryptophan and tyrosine residues. Inflection points of the measured curve represent the melting temperature T_{M} (Alexander *et al.*, 2014). Owing to low sample consumption, the technique can be used for high-throughput screening of protein stability in many different storage buffers.

Dynamic light scattering measures the size distribution profile of molecules in solution. Monochromatic polarized light (usually laser) is scattered on a sample and detected after going through a polarizer. Fluctuations in scattered light intensity due to diffusing particles depend on the particle size (Berne and Pecora, 2000; Stetefeld *et al.*, 2016). Hence, the method can indicate aggregation of a sample and allows high throughput screening of storage buffers while monitoring protein stability over time (Meyer *et al.*, 2015).

Circular dichroism spectroscopy investigates the secondary structure of proteins. The principle of this method is the measurement of differential absorption of left- and right-handed circularly polarized light in the UV range (Greenfield, 2006; Whitmore and Wallace, 2008). The reported degrees of ellipticity are defined as

$$\tan\theta = \frac{E_L - E_R}{E_L + E_R}, \tag{26}$$

where E_L and E_R denote electric field vectors magnitudes of the left- and right-circularly polarized light, respectively. α-helices, β-sheets and some other conformations in a protein structure have characteristic spectral signatures. Thus, circular dichroism spectra reveal the fraction of particular secondary-structure elements of a macromolecule.

1.5.2 Crystallization

To solve an atomic structure using the diffraction methods, a studied macromolecule is crystallized to form ideally a three-dimensional single crystal. A small amount of a

purified protein sample is mixed with a precipitant solution that usually consists of a buffer, salts, polymers (*e.g.* polyethylene glycol), organic solvents and possibly also of other molecules (Ducruix and Giegé, 1992). The used protein concentration is high to exceed its solubility limit and create a supersaturated state. To reach the equilibrium, the protein molecules precipitate which creates crystallization nuclei and these nuclei further grow and form crystals.

Crystallization is a delicate process that depends on several factors, mainly on the concentration of individual compounds, pH and temperature. However, it is not usually possible to predict the optimal content of a precipitant solution. Thus, the first step is a high-throughput screening of numerous different conditions, often automated with a crystallization robot (Bergfors, 2007). The initial hits are then optimized by the modification of initial parameters.

Apart from microbatch and dialysis, vapor diffusion is the most commonly used technique for crystallization. The hermetically sealed system for the vapor diffusion method consists of a high-volume reservoir, containing the precipitant solution, and a small-volume drop from a protein sample mixed with the precipitant solution (Doublié, 2007). The drop can be in the hanging or sitting setup (Figure 5). The precipitant concentration is lower in the drop in comparison with the reservoir. This causes the water diffusion from the drop and a subsequent desired increase in the concentration of protein and precipitant in the drop.

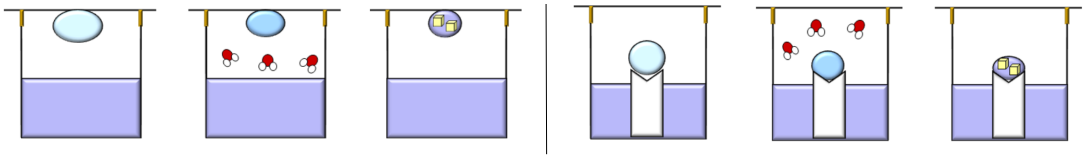


Figure 5: Crystallization process using the vapor diffusion method in the hanging drop (a) and sitting drop (b) setup (Russo Krauss *et al.*, 2013).

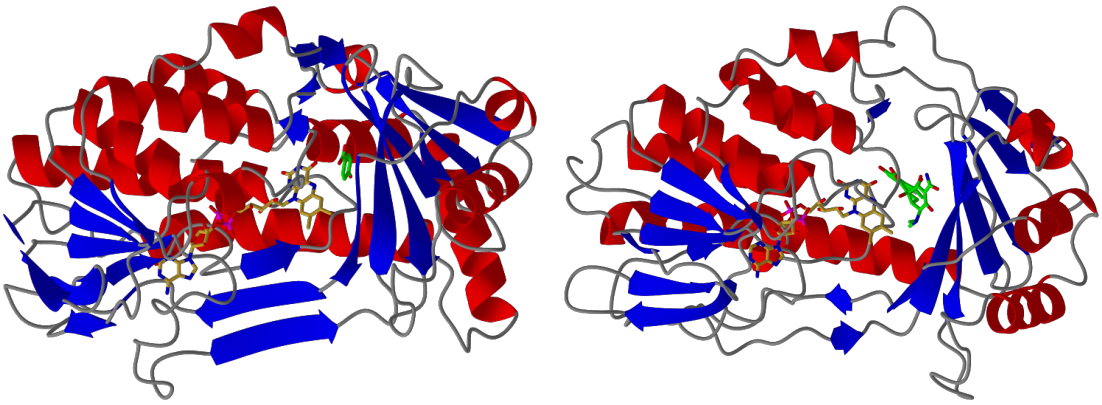
1.6 FAD-dependent monooxygenases

FAD-dependent monooxygenases (FDO) from the enzyme classes EC 1.13 and EC 1.14 represent a wide group of flavoenzymes, which is further classified into eight subclasses (Paul *et al.*, 2021). The enzymes in the class A FDO meet the following characteristics (van Berkel *et al.*, 2006):

- The structure contains an FAD-binding domain that binds tightly the FAD prosthetic group, especially the dinucleotide moiety using a Rossmann fold.
- Dependency on the nicotinamide adenine dinucleotide (NADH) or nicotinamide adenine dinucleotide phosphate (NADPH) cofactors that are released immediately after their oxidation to NAD+ or NADP+.

The cofactor NAD(P)H is needed for the reduction of FAD to $FADH_2$ that subsequently reacts with molecular oxygen to form a reactive adduct

C4a-hydroperoxyflavin (Crozier-Reabe and Moran, 2012). The hydroperoxy group then carries out an electrophilic attack on an aromatic ring of a substrate (Huijbers *et al.*, 2014). Thus, the typical substrates are aromatic molecules that contain an amino or hydroxyl group. The substrate specificity of individual enzymes is usually narrow. The comprehensively investigated prototype enzyme of the class A FDO represents 4-hydroxybenzoate 3-monooxygenase (EC 1.14.13.2) (Entsch and Van Berkel, 1995). Its structure is shown in Figure 6a.



(a) (b)

Figure 6: Crystal structures of representatives of the class A FAD-dependent monooxygenases in secondary structure representation. α-helices are colored in red and β-sheets in blue. The prosthetic group FAD is shown as yellow sticks and the substrates bound in the active site as green sticks. (a) 4-hydroxybenzoate 3-monooxygenase from *Pseudomonas fluorescens* in complex with *p*-hydroxybenzoic acid (Schreuder et al. 1989; PDB entry 1PBE). (b) Tetracycline destructase TetX from *Bacteroides thetaiotaomicron* in complex with chlortetracycline (Volkers *et al.*, 2011; PDB entry 2Y6R). The graphics were created with *CCP4mg* (McNicholas *et al.*, 2011).

1.6.1 Tetracycline destructases

The family of tetracycline destructases belongs also to the class A FDO. They are responsible for one of the resistance mechanism of bacteria against antibiotics. These enzymes are capable of the covalent modifications of tetracycline antibiotics which cause the destruction of the antibiotic scaffold (Markley and Wencewicz, 2018; Yang *et al.*, 2004).

Tetracycline antibiotics are a group of compounds that exhibit bacteriostatic activity against a wide range of microorganisms (Chopra and Roberts, 2001). Their scaffold is composed of four rings (Figure 7). They bind to the 30S bacterial ribosome subunit which restricts the proteosynthesis in a pathogen cell (Brodersen *et al.*, 2000). Thus, they play an important role in various therapies. Nevertheless, the application of the original tetracyclines variants (*e.g.* tetracycline, oxytetracycline) is nowadays limited, basically due to the onset of antibiotic resistance (Grossman, 2016; Nguyen *et al.*, 2014).

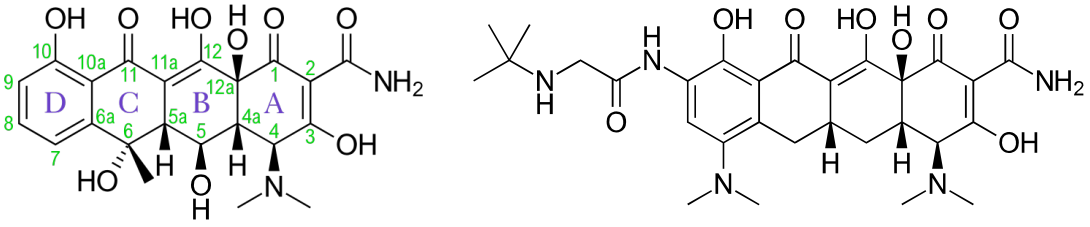


Figure 7: (a) Oxytetracycline (Chopra and Roberts, 2001). (b) Tigecycline.

There are five known molecular mechanisms of microbial resistance to tetracycline antibiotics: efflux, ribosome protection, ribosome mutation, reduced drug permeability, and enzymatic inactivation. The most common types are efflux pumps and ribosome protection proteins (Nguyen *et al.*, 2014). Fortunately, these both resistance mechanisms were overcome by exploration of the third and fourth generation of synthetic tetracycline derivatives, *e.g.* tigecycline (Jenner *et al.*, 2013), eravacycline (Zhanel *et al.*, 2020) and omadacycline (Tanaka *et al.*, 2016). Despite a lower prevalence, a current emerging problem with a devastating medical impact is the enzymatic inactivation (Markley and Wencewicz, 2018). In the particular case of tetracyclines, the enzymatic inactivation is caused by the tetracycline destructases, as mentioned above. This enzymatic family of tetracycline destructases (EC 1.14.13.231) can be classified into two groups: the TetX variants (TetX1~14) (Cheng *et al.*, 2022) and Tet(47-55) (Forsberg *et al.*, 2015) which possess at most 24.4% sequence identity to TetX (Markley and Wencewicz, 2018).

The enzymatic inactivation of tetracycline antibiotics was reported quite recently (Moore *et al.*, 2005; Yang *et al.*, 2004). This discovery was followed by the solution of crystal structures of tetracycline destructases TetX2 (Walkiewicz *et al.*, 2011) and TetX (Volkers *et al.*, 2010, 2011, 2013) (Figure 6b) (both from *Bacteroides thetaiotaomicron*), the latter also in complex with chlortetracycline, iodtetracycline, minocycline and tigecycline. In 2015, a subgroup of tetracycline destructases was identified and functionally characterized (Forsberg *et al.*, 2015) and subsequently, the corresponding crystal structures of Tet(50), Tet(51), Tet(55) and Tet(56) were solved (Park *et al.*, 2017). The structure of Tet(50) revealed FAD in two different conformations, denoted as IN and OUT, and a novel binding mode of tetracycline molecules in the active site. Recently, crystal structures of TetX7 (Gasparrini *et al.*, 2020) and TetX4 (Cheng *et al.*, 2021) are also available. The crystal structures provided the basis for the understanding of which residues are crucial for the catalytic activity (Cui *et al.*, 2021), for the design of novel inhibitors (Markley *et al.*, 2019) and for other related studies (Cheng *et al.*, 2022; Liu *et al.*, 2020; Wang *et al.*, 2019).

The protein structure is composed of an FAD-binding domain, a substrate binding domain and one (TetX variants, Figure 6b) or two (Tet(47-56)) C-terminal helices. The NADPH cofactor is required for the catalytic activity. The complex structures reveal the different spatial orientations of the bound tetracycline molecules in the active site of TetX and Tet(50). The binding mode determines which carbon atom of

a substrate is directly accessible to the reactive C4a-hydroperoxyflavin and is then hydroxylated. The potential oxidative sites are C11a-enol center for TetX, C12-carbonyl center for Tet(56) and C1-carbonyl, C2-enol and C3-carbon centers for Tet(50) and Tet(55). The tetracycline products are usually unstable and further degrade. The tetracycline destructases can be inhibited with anhydrotetracycline (Park *et al.*, 2017) and its semisynthetic analogs which are under investigation, *e.g.* anhydrodemeclocycline (Markley *et al.*, 2019).

Aims of the study

The thesis deals with the following two major aims of the study:

- Development of the scientific software tool *PAIREF* providing automation of the paired refinement protocol to determine an optimal high-resolution cutoff of diffraction data.
  Described in Part 2)

- Expression, purification and characterization of the tetracycline-modifying enzyme from *Stenotrophomonas maltophilia* (*Sm*TetX, NCBI Reference Sequence WP_049406473). Testing of its enzymatic activity towards a tetracycline representative. Crystallization and structural analysis including the application of the developed *PAIREF* program during the data processing.
  Described in Part 3)

Part 2)

Development of methods

2.1 Journal article A)
Paired refinement under the control of *PAIREF*

In this scientific article, the software tool *PAIREF* providing automation of paired refinement is introduced to the scientific community. This procedure allows the determination of an optimal high-resolution cutoff owing to the linking of quality of a structure model and diffraction data. Our article follows the previous works that originally described the paired refinement protocol (Diederichs and Karplus, 2013; Karplus and Diederichs, 2012, 2015).

In this manuscript, a general overview of the procedure is reported. We describe in detail the *PAIREF* algorithm and technical aspects of implementation. Subsequently, the application of the program is thoroughly demonstrated on six examples including the cysteine dioxygenase dataset investigated in the first work (Karplus and Diederichs, 2012) and a generated artificial lysozyme dataset. The impact of the quality of an input structure model was also investigated.

We show that paired refinement under defined conditions produces structure models closest to the truth. Many structure models can be improved by inclusion of data beyond conservative cutoffs.

Author contribution

The author developed the whole code of the program *PAIREF* and wrote the documentation and web pages of the project (*https://pairef.fjfi.cvut.cz*). The author designed and carried out numerous paired refinement runs, interpreted data and was responsible for writing the manuscript.

IUCrJ
ISSN 2052-2525
BIOLOGY | MEDICINE

Paired refinement under the control of *PAIREF*

Martin Malý,[a,b] Kay Diederichs,[c] Jan Dohnálek[b] and Petr Kolenko[a,b]*

[a]Faculty of Nuclear Sciences and Physical Engineering, Czech Technical University in Prague, Břehová 7, Prague 11519, Czech Republic, [b]Institute of Biotechnology of the Czech Academy of Sciences, Biocev, Průmyslová 595, Vestec 25250, Czech Republic, and [c]University of Konstanz, Box M647, Konstanz 78457, Germany. *Correspondence e-mail: petr.kolenko@fjfi.cvut.cz

Received 13 March 2020
Accepted 29 April 2020

Edited by K. Moffat, University of Chicago, USA

Keywords: macromolecular crystallography; *PAIREF*; X-ray diffraction; paired refinement; high-resolution limit.

Supporting information: this article has supporting information at www.iucrj.org

Crystallographic resolution is a key characteristic of diffraction data and represents one of the first decisions an experimenter has to make in data evaluation. Conservative approaches to the high-resolution cutoff determination are based on a number of criteria applied to the processed X-ray diffraction data only. However, high-resolution data that are weaker than arbitrary cutoffs can still result in the improvement of electron-density maps and refined structure models. Therefore, the impact of reflections from resolution shells higher than those previously used in conservative structure refinement should be analysed by the paired refinement protocol. For this purpose, a tool called *PAIREF* was developed to provide automation of this protocol. As a new feature, a complete cross-validation procedure has also been implemented. Here, the design, usage and control of the program are described, and its application is demonstrated on six data sets. The results prove that the inclusion of high-resolution data beyond the conventional criteria can lead to more accurate structure models.

1. Introduction

Crystallographic resolution is understood as the minimum plane spacing given by Bragg's law for a particular set of X-ray diffraction intensities that are included in the structure analysis (*Online Dictionary of Crystallography*, https://dictionary.iucr.org/Resolution). In contrast, optical resolution is defined as the expected minimum distance between two resolved peaks in the electron-density map (Vaguine *et al.*, 1999). The resolution of data is limited due to a decrease in the intensity-to-noise ratio of reflections with the resolution. The weakness of the high-resolution data is caused by several factors, including the Lorentz-polarization factor, temperature factor and crystal imperfection. Therefore, the diffraction data are usually cut off at a certain resolution, with the aim of rejecting the data that do not improve the model.

In previous decades, conservative criteria were applied to estimate the resolution of crystallographic data. These criteria were based on a user-defined value of data quality indicators such as the signal-to-noise ratio $\langle I/\sigma(I)\rangle$, the disagreement residual of multiple observations $R_{\rm merge}$, *etc.* (Evans, 2011). Later, the Pearson correlation coefficient $CC_{1/2}$, quantifying the internal consistency of observations, was added to these criteria (Karplus & Diederichs, 2012). Inspection of the data deposited in the PDB (Berman *et al.*, 2000) shows that there is no consensus in the application of these statistics. Moreover, the possibility of improvement of a refined model by employing a different resolution range was often not considered. Nowadays, the application of strict cutoff values on selected data quality indicators has been shown to be an obsolete approach (Diederichs & Karplus, 2013; Evans &

Murshudov, 2013). Very recently, it became possible to estimate the information gain from each reflection using likelihood-based methods (Read *et al.*, 2020). Yet this approach does not answer the question of which high-resolution cutoff should be used with current refinement programs.

The ambiguity in the high-resolution-cutoff estimation has been removed with the advent of the 'paired refinement' protocol (Karplus & Diederichs, 2012). Initially, a conservative criterion is applied as usual to the high-resolution data and the phase problem is solved. Usually, the model is then significantly improved by refinement. In the paired refinement protocol, the influence of the previously rejected high-resolution data during the structure refinement is tested. The structure model is refined stepwise against data at higher and higher resolution until no improvement of the model is observed. More specifically, each increase in resolution is checked against the original resolution for its added value, particularly by comparing R values of models against the same data. Only those resolution shells that prove beneficial are included in the final data set, against which the structure is refined.

In this paper, we present a new tool – *PAIREF* – which helps to make the decision about the useful resolution of the data set. The program performs paired refinement for validation of the high-resolution data in a fully automatic way. *PAIREF* is not the first utility that implements paired refinement since a similar function is present in *PDB-REDO* (Joosten *et al.*, 2014). Nevertheless, *PAIREF* provides additional features (*e.g.* complete cross-validation, modification of the structure refinement protocol) and reports that naturally require more extensive input, and allows a user to make a more sophisticated decision.

2. Design and implementation

PAIREF is a command-line tool that can be installed as a module into the *CCTBX* toolbox (Grosse-Kunstleve *et al.*, 2002) on various platforms (GNU/Linux, MS Windows). Currently, it has been developed in Python 2.7 (Hunter, 2007; Rossum, 1995) but is ready to move to Python 3. It depends on the following programs of the *CCP*4 software package (Winn *et al.*, 2011): *REFMAC*5 (Murshudov *et al.*, 2011), *SFCHECK* (Vaguine *et al.*, 1999), *MTZDUMP*, *SFTOOLS* and *BAVERAGE*; and on the module *pdbtools* (Adams *et al.*, 2010) from *CCTBX*. Input parameters can be specified in order to place the protocol under the full control of the user.

A typical command-line example for a *PAIREF* job is `cctbx.python -m pairef --XYZIN starting_model_2-4A.pdb --HKLIN data_2A.mtz --HKLIN_UNMERGED data_2A_unmerged.mtz -i 2.4 -r 2.3,2.2,2.1,2.0`, which executes refinements of the structure model `starting_model_2-4A.pdb` (previously refined at 2.4 Å) for a series of cutoffs (stepwise 2.3, 2.2, 2.1 and 2.0 Å). Specification of unmerged data (MTZ, unmerged *Scalepack* or *XDS*/*XSCALE* file types) is only required if comparison of CC_{work} with CC* (see below) should be enabled.

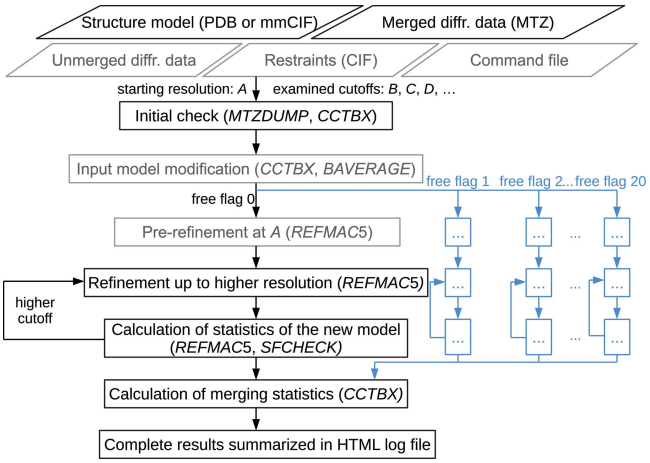


Figure 1
Schematic diagram of the *PAIREF* algorithm. Optional input files and routines are drawn in grey, the complete cross-validation protocol is outlined in blue.

2.1. Parameters and algorithm

The algorithm implemented in *PAIREF* depends on the amount of data provided by the user. The minimal function of the program requires the following input files: structure model refined at the starting resolution (PDB or mmCIF format) and higher-resolution merged diffraction data in MTZ format which have the same free reflection flags as the data previously used in the refinement (Fig. 1). Nevertheless, the minimal requirement is not sufficient for deep data analysis including statistics such as CC*, *etc.* The protocol can be further supplemented by the full-resolution unmerged data for calculating merging statistics, by the external restraints in CIF format in the case where non-standard ligands are present and by the command file for *REFMAC*5 (alternatively generated by *PDB-REDO*) for better control of the structure refinement. Moreover, a definition of domains for translation–libration–screw (TLS) refinement can be provided by the user. The program allows the selection of resolution shells (with a default width of 0.05 Å) and optional model modifications before the paired refinement.

Our paired refinement protocol with *REFMAC*5 is an adaptation of the original protocol that has been performed with *phenix.refine* (Karplus & Diederichs, 2012; Afonine *et al.*, 2012). Initially, the input files are checked using *MTZDUMP* and *CCTBX* for consistency. The model is then refined against the data up to resolution B (higher than A), and this model is compared with the original one – both against the data at resolution A (see Section 2.2). This step is then repeated from resolution B up to resolution C (higher than B) and reproduced again until the maximum limit is reached. CC_{work} and CC_{free} statistics are calculated using *SFTOOLS* (Karplus & Diederichs, 2012). Finally, merging statistics are calculated using the *CCTBX* library if unmerged diffraction data were provided.

As an option, *PAIREF* provides a complete cross-validation protocol (Brünger, 1993; Jiang & Brünger, 1994) – also referred to as k-fold cross-validation (Luebben & Gruene,

2015) – to investigate the impact of the selection of free reflections. Here, the paired refinement protocol is run in parallel for each selection individually. To remove the bias given by previous refinement with a particular set of free reflections, a number of optional input model modifications prior to refinement have been implemented: the perturbation of the atomic coordinates, the reset of atomic displacement parameters (ADPs) to a particular or average value and the addition of a fixed value to them (achieved by module *pdbtools* from *CCTBX* and *BAVERAGE*). In the final report, both the averaged statistics as well as the individual statistics for each selection are reported. Application of this protocol is demonstrated on a data set from cysteine dioxygenase (Section 3.3). The complete cross-validation requires the *CCP*4-style test set description in the input MTZ file, *i.e.* multiple free reflection labels must be present.

The program *PAIREF* does not have any decision-making routines and it remains up to the user to decide on the resolution cutoff based on the comprehensive analysis that was performed. Structure refinement is a multiparametric calculation and the user should be aware of potential problems. For example, nonconvergent refinement may result in misleading statistics and a suboptimal model (Tickle, 2011). One of the parameters that may potentially play a role is the FFT grid size (Drenth & Jeroen, 2010).

2.2. Program output and interpretation of results

Paired refinement does not reduce the problem of high-resolution cutoff estimation to a single monitoring statistic. Rather, a comprehensive data analysis is summarized on an HTML page. Here, various plots, tables and links to many intermediate files and log files are presented or easily accessible via hyperlinks.

The first monitoring statistics reported by *PAIREF* are the differences in R values between the models refined at adjacent resolutions (both computed at the lower resolution to provide a valid comparison). A decrease in $R_{\rm free}$ is expected in shells beneficial to the model quality. However, a constant $R_{\rm free}$ and a simultaneous increase in $R_{\rm work}$ are usually acceptable as well because these indicate less overfitting of the structure model (Karplus & Diederichs, 2012). Therefore, the next monitoring statistic is $R_{\rm gap}$ ($R_{\rm gap} = R_{\rm free} - R_{\rm work}$) which is calculated at the starting resolution (corresponding to resolution A in Section 2.1) for all analyzed shells. This is an implementation of a previously published protocol (Winter *et al.*, 2018). In the case of the complete cross-validation protocol, R values for each set of free reflections and average values are reported. Moreover, the standard deviations of R values of structure models refined using different free reflection sets are calculated (Kleywegt & Brünger, 1996).

However, the overall R values are not the only parameters to be taken into account when deciding on the high-resolution cutoff. The analysis is further supplemented by plots of $R_{\rm work}$, $R_{\rm free}$, $\mathrm{CC}_{\rm work}$ and $\mathrm{CC}_{\rm free}$ ($\mathrm{CC}_{\rm work}$ and $\mathrm{CC}_{\rm free}$ are correlation coefficients between experimental and calculated intensities) of the refined structure models at defined resolution. Since a perfect model gives an R value of 0.42 against random data (*i.e.* pure noise) – assuming non-tNCS (translational non-crystallographic symmetry) data from a non-twinned crystal (Evans & Murshudov, 2013) – a higher R value in the (current) high-resolution shell indicates either the involvement of high-resolution data without information content (the data are even worse than noise), or poor quality of the model, or the presence of tNCS.

When unmerged data are available, values of CC* are added to the $\mathrm{CC}_{\rm work}$ and $\mathrm{CC}_{\rm free}$ plots. Comparison of CC values (correlation coefficients) with CC* serves for direct linking of the data and structure model quality (Diederichs & Karplus, 2013; Karplus & Diederichs, 2012). $\mathrm{CC}_{\rm work}$ or $\mathrm{CC}_{\rm free}$ greater than CC* in a high-resolution shell indicates undesirable overfitting of the structure model as the calculated intensities agree with the observed data better than the (usually unavailable) true data. Owing to the independence of CC* on a model, its comparison with $\mathrm{CC}_{\rm work}$ is just as informative as comparison with $\mathrm{CC}_{\rm free}$. However, the usage of $\mathrm{CC}_{\rm work}$ should be preferred since it is based on much more data.

For additional information, *PAIREF* reports the optical resolution as calculated using *SFCHECK* for each resolution cutoff. When all previous procedures are finished and unmerged diffraction data are available, the merging statistics are listed in a table and shown in graphs. Finally, the progress of the refinement procedures is reported to check for convergence *etc.*

2.3. Distribution and documentation

Full documentation of *PAIREF* is available online at https://pairef.fjfi.cvut.cz and the program is distributed at https://pypi.org/project/pairef/.

3. Examples

The functionality and versatility of *PAIREF* have been thoroughly tested on a number of cases. Here, we selected six structures and data sets to demonstrate the broad application potential of the tool: simulated data for lysozyme from *Gallus gallus* (SIM) (Holton *et al.*, 2014), and measured data for thermolysin from *Bacillus thermoproteolyticus* (TL) (Winter *et al.*, 2018), a cysteine-bound complex of cysteine dioxygenase from *Rattus norvegicus* (CDO) (Karplus & Diederichs, 2012), endothiapepsin from *Cryphonectria parasitica* in complex with fragment B53 (EP) (Huschmann *et al.*, 2016), interferon gamma from *Paralichthys olivaceus* (POLI) (Zahradník *et al.*, 2018) and bilirubin oxidase from *Myrothecium verrucaria* (BO) (Koval' *et al.*, 2019). All the results are available from http://doi.org/10.5281/zenodo.3687267.

A comprehensive summary of crystallographic data as well as the refinement statistics are shown in Tables 1 and 2. To be consistent with the previous results, the free reflection flags from the original data were preserved except for TL, because of inaccessibility.

Table 1
Data collection and merging statistics.

Values for the highest resolution shell in the case of conservative cutoff are given in parentheses () and for the cutoff chosen as optimal are given in square brackets []. SIM represents a simulated data set generated by *MLFSOM* (Holton *et al.*, 2014).

Data set	SIM	TL	CDO	EP	POLI	BO†
Data set DOI	10.15785/SBGRID/746	10.5281/zenodo.49559	10.15785/SBGRID/751	10.18430/m34y4g	10.5281/zenodo.3369718	10.18430/m36i3j
X-ray source	*MLFSOM*	BL I03, Diamond Light Source	BL 5.0.1. Advanced Light Source	BL14.1, BESSY II	BL14.1, BESSY II	BL14.1, BESSY II
Wavelength (Å)	1.0000	1.2276	0.9774	0.9184	0.9184	0.9184
Detector	Simulated PILATUS 6M	PILATUS 6M	ADSC	PILATUS 6M	PILATUS 6M	MAR mosaic CCD
Temperature (K)	N/A	N/A	100	100	100	100
Crystal-to-detector distance (mm)	150.0	209.4	150.0	180.8	446.3	313.5
Oscillation angle/range (°)	1/100	0.1/720	1/218	0.1/200	0.1/360	0.5/108.5
Resolution range (Å)	38.64–1.30 (1.98–1.72) [1.40–1.30]	79.98–1.50 (1.90–1.80) [1.60–1.50]	41.96–1.50 (2.10–2.00) [1.60–1.50]	49.64–1.20 (1.51–1.44) [1.25–1.20]	47.32–2.00 (2.38–2.30) [2.10–2.00]	47.35–2.59 (2.67–2.59) ⟨2.59–2.50⟩†
Space group	$P4_32_12$	$P6_122$	$P4_32_12$	$P2_1$	$P2_12_12_1$	$F222$
Unit-cell parameters						
a (Å)	77.24	92.35	57.63	45.20	58.27	134.5
b (Å)	77.24	92.35	57.63	73.10	79.76	204.1
c (Å)	38.66	127.71	122.39	52.57	94.64	227.0
α (°)	90.00	90.00	90.00	90.00	90.00	90.00
β (°)	90.00	90.00	90.00	109.25	90.00	90.00
γ (°)	90.00	120.00	90.00	90.00	90.00	90.00
Wilson B factor (Å^2)	12.2	20.3	24.2	15.5	64.5	44.5
No. reflections	166742 (30516) [16791]	3714005 (341691) [510558]	522379 (33610) [60331]	371954 (29343) [42646]	393534 (23854) [49772]	399548 (27236) ⟨27420⟩†
No. unique reflections	28932 (4336) [5345]	50760 (4391) [8252]	33898 (1938) [5862]	97408 (7460) [10944]	30377 (1928) [4021]	48468 (4177) ⟨5353⟩†
No. additional unique reflections‡	16029 {1.72–1.30}	20518 {1.80–1.50}	25117 {2.00–1.50}	40250 {1.44–1.20}	10202 {2.30–2.00}	0
Multiplicity	5.8 (7.0) [3.1]	73.2 (77.8) [61.9]	15.4 (17.3) [10.3]	3.8 (3.9) [3.9]	13.0 (12.4) [12.4]	8.2 (6.5) ⟨5.1⟩†
Completeness (%)	98.6 (99.9) [93.4]	97.6 (98.3) [91.8]	100.0 (100.0) [100.0]	96.8 (96.3) [94.6]	99.7 (100.0) [98.5]	100.0 (100.0) ⟨99.8⟩†
Mean $I/\sigma(I)$	5.9 (4.0) [0.3]	13.3 (4.4) [0.8]	22.7 (18.1) [0.9]	6.6 (1.7) [0.5]	9.0 (0.9) [0.1]	13.8 (1.7) ⟨1.2⟩†
R_{meas}	0.131 (0.254) [2.233]	0.223 (1.143) [4.828]	0.150 (0.334) [2.133]	0.117 (0.777) [2.500]	0.154 (2.907) [17.721]	0.150 (1.143) ⟨1.338⟩†
R_{pim}	0.052 (0.094) [1.153]	0.025 (0.127) [0.598]	0.037 (0.079) [0.654]	0.059 (0.385) [1.247]	0.043 (0.816) [4.963]	0.052 (0.445) ⟨0.584⟩†
$CC_{1/2}$	0.992 (0.971) [0.179]	1.000 (0.961) [0.445]	0.999 (0.996) [0.437]	0.998 (0.694) [0.225]	0.999 (0.578) [0.027]	0.997 (0.652) ⟨0.524⟩†
Resolution range (Å)§	38.64–1.35	79.98–1.43	41.96–1.42	49.64–1.11	47.32–1.90	47.35–2.30
CC*	0.998 (0.993) [0.551]	1.000 (0.990) [0.785]	1.000 (0.999) [0.780]	0.999 (0.905) [0.606]	1.000 (0.856) [0.229]	0.999 (0.888) ⟨0.829⟩†

† For the BO data set, values for a resolution shell beyond the optimal cutoff are listed in angled brackets ⟨⟩. ‡ Number of additional reflections suggested by paired refinement results to be involved in the refinement in contrast to the starting resolution. Added resolution range, in Å, is given in {} brackets. § Range where $CC_{1/2}$ is significantly different from 0 at the 1:1000 level.

3.1. Simulated data set of lysozyme

The ability to generate artificial X-ray diffraction patterns based on a well defined 'true' structure offers the possibility of monitoring the progress of paired refinement, especially the convergence of the refined models towards the 'true' structure.

We generated one hundred diffraction images using a modified structure of lysozyme (data set SIM). At first, all alternative conformations were removed from the structure with the PDB entry 1h87 (originally determined at 1.72 Å resolution) (Girard *et al.*, 2002). The data collection was simulated using *MLFSOM* (Holton *et al.*, 2014) with a crystal-to-detector distance of 150 mm. *MLFSOM* also simulated global radiation damage for a beam of 8.4×10^{10} photons s^{-1} and 100 µm diameter, exposure of 0.1 s and a crystal size of 77.8 µm. Afterwards, the diffraction data set was processed using *DIALS*/*AIMLESS* (Evans & Murshudov, 2013; Winter *et al.*, 2018) or *XDS*/*XSCALE* (Kabsch, 2010) up to a resolution of 1.20 Å, although the $CC_{1/2}$ values become not significantly different from zero (at the 1:1000 level) at 1.35 Å resolution.

The input model for paired refinement was generated from the structure used for the generation of the diffraction images by perturbation of atomic coordinates by an average of 0.25 Å; the ADPs were set to their mean value (15 Å^2). In the final preparation step, several cycles of restrained refinement at the starting resolution (1.72 Å) against the processed simulated data were performed. In the next step, we performed the paired refinement protocol using *PAIREF*.

Structure models refined against the simulated data set have considerably lower R values when compared with the other structures (based on real experimental data) mentioned later (R_{free}= 0.071 for SIM versus R_{free}= 0.195 for TL, both at

Table 2
Structure refinement and validation statistics.

Values are listed for the models refined at the starting and the optimal resolution in square brackets []. ΔR is the difference between R values relating to the model refined at the optimal and the starting resolution (both calculated at the starting resolution). SIM is a simulated data set generated by *MLFSOM* (Holton *et al.*, 2014).

Data set	SIM	TL	CDO†	EP	POLI	BO‡
Resolution range (Å)	38.64–1.72 [38.64–1.30]	79.98–1.80 [79.98–1.50]	41.96–2.00 [41.96–1.50]	49.64–1.44 [49.64–1.20]	47.32–2.30 [47.32–2.00]	47.35–2.59 ⟨47.35–2.50⟩‡
Optical resolution (Å)	1.41 [1.25]	1.52 [1.42]	1.50 [1.30]	1.30 [1.15]	2.16 [2.08]	2.01 ⟨1.99⟩‡
R_{work}	0.0605 [0.1047]	0.1580 [0.1742]	0.1560 (σ = 0.0010) [0.2070 (σ = 0.0010)]	0.2017 [0.2241]	0.2236 [0.2412]	0.1754 ⟨0.1881⟩‡
ΔR_{work}	−0.0011	0.0028	0.0048	0.0026	−0.0003	0.0002
R_{free}	0.0711 [0.1112]	0.1954 [0.2037]	0.2060 (σ = 0.0080) [0.2380 (σ = 0.0070)]	0.2566 [0.2656]	0.2972 [0.3152]	0.2408 ⟨0.2498⟩‡
ΔR_{free}	−0.0042	−0.0023	−0.0090	−0.0051	−0.0016	0.0003
CC_{work}	0.9822 [0.9826]	0.9615 [0.9630]	0.9590 (σ = 0.0020) [0.9650 (σ = 0.0010)]	0.9436 [0.9306]	0.9199 [0.9387]	0.9450 ⟨0.9471⟩‡
CC_{free}	0.9915 [0.9920]	0.9467 [0.9498]	0.9400 (σ = 0.0200) [0.9500 (σ = 0.0100)]	0.9177 [0.9069]	0.8678 [0.8690]	0.9151 ⟨0.9168⟩‡
Average ADP (Å^2)	13.67 [13.59]	22.55 [23.43]	14.47 [19.17]	13.10 [12.76]	70.09 [68.17]	45.10 ⟨46.87⟩‡
RMSD bond lengths (Å)	0.012 [0.013]	0.012 [0.012]	0.011 [0.013]	0.017 [0.014]	0.012 [0.013]	0.008 ⟨0.008⟩‡
RMSD bond angles (°)	1.915 [1.942]	1.649 [1.707]	1.739 [1.853]	1.846 [1.797]	1.829 [2.005]	1.326 ⟨1.654⟩‡
No. of non-hydrogen atoms	1217	2816	1836	2459	2286	9511
Ramachandran: favoured (%)	91.3 [92.1]	93.6 [96.6]	97.3 [97.3]	97.4 [97.0]	93.1 [94.2]	90.8 ⟨90.7⟩‡
Ramachandran: outliers (%)	0.0 [0.0]	1.0 [1.0]	0.0 [0.0]	0.0 [0.3]	1.5 [1.5]	1.4 ⟨1.4⟩‡

† In the case of complete cross-validation (data set CDO), R values and CC values averaged over all 20 free reflection sets and the associated standard deviation σ are listed. The remaining statistics relate to the refinements with free reflection set 0. ‡ For the BO data set, values for a resolution shell beyond the optimal cutoff are listed in angled brackets ⟨⟩.

1.72 Å). This effect, caused by the simulated character of the data, was also observed in the original work by Holton *et al.* (2014). However, the trends of nearly all indicators of data quality are similar to those of the real cases [see Fig. 2(*a*)]. Based on the plot of stepwise differences in overall R values, we decided to estimate the high-resolution limit as 1.3 Å because the R values increase for resolution shells beyond that limit.

We monitored the root-mean-square deviation (RMSD) values (DeLano Scientific, 2017) calculated on all 1217 atoms of the simulated structure with respect to the original structure model [Fig. 2(*c*)]. A systematic decrease was observed for the atomic coordinates when reflections from an additional high-resolution shell were added to the refinement up to 1.3 Å resolution. This is in agreement with the high-resolution cutoff based on the differences in overall R-values behaviour only. In general, the RMSD of ADP values calculated for all the atoms (see equation given in the supporting information) follow a similar but not identical trend. Moreover, they continue to decrease and converge to the 'true' value even for the highest resolution shell which was later omitted from the data based on the other data quality indicators. As a result of our calculations, we suggest here application of a high-resolution cutoff at 1.3 Å when using our combination of programs and following our refinement protocol. Similar results were also obtained using *XDS*/*XSCALE* for data processing.

3.2. Thermolysin

Successful application of paired refinement was previously demonstrated on the crystal structure of thermolysin (TL) from *B. thermoproteolyticus* (Winter *et al.*, 2018). In the original protocol, the structure was modified (perturbation of atomic positions) and refined at a defined high-resolution limit in the range from 1.80 to 1.50 Å. Model improvement was monitored on R_{gap} only, which decreased until 1.56 Å resolution. A further increase in the resolution did not cause a substantial change of R_{gap}.

To reproduce most of the original procedures by Winter *et al.*, the diffraction data were processed with *xia2* (Winter, 2010) using *DIALS*/*AIMLESS* software. The structure of thermolysin (PDB entry 3n21; Behnen *et al.*, 2012) was used as a starting model. The atomic coordinates were perturbed and all ADPs were generally set to their average value of 22 Å^2 with *phenix.pdbtools* (Adams *et al.*, 2010). A total of 30 cycles of restrained refinement were performed with *REFMAC*5 at a resolution of 1.80 Å. After that, ligands (peptide in the active site, three molecules of DMSO) and solvent were built in *Coot* (Emsley *et al.*, 2010), refined with *REFMAC*5 and finally used in *PAIREF* to analyse the high-resolution cutoff.

We performed two *PAIREF* runs that added stepwise high-resolution shells with a width of 0.10 and 0.01 Å. R_{free} has a decreasing trend up to 1.50 Å for the first run [Fig. 2(*d*)], which suggests that the data should be cut at this resolution. Moreover, the plot of R_{gap} [Fig. 2(*f*)] from the second run further confirms a good agreement between the previously published results and our calculations.

3.3. Cysteine dioxygenase

The cysteine-bound complex of cysteine dioxygenase from *R. norvegicus* (CDO) (Simmons *et al.*, 2008) was the first macromolecular crystal structure on which the paired refinement protocol was demonstrated (Karplus & Diederichs,

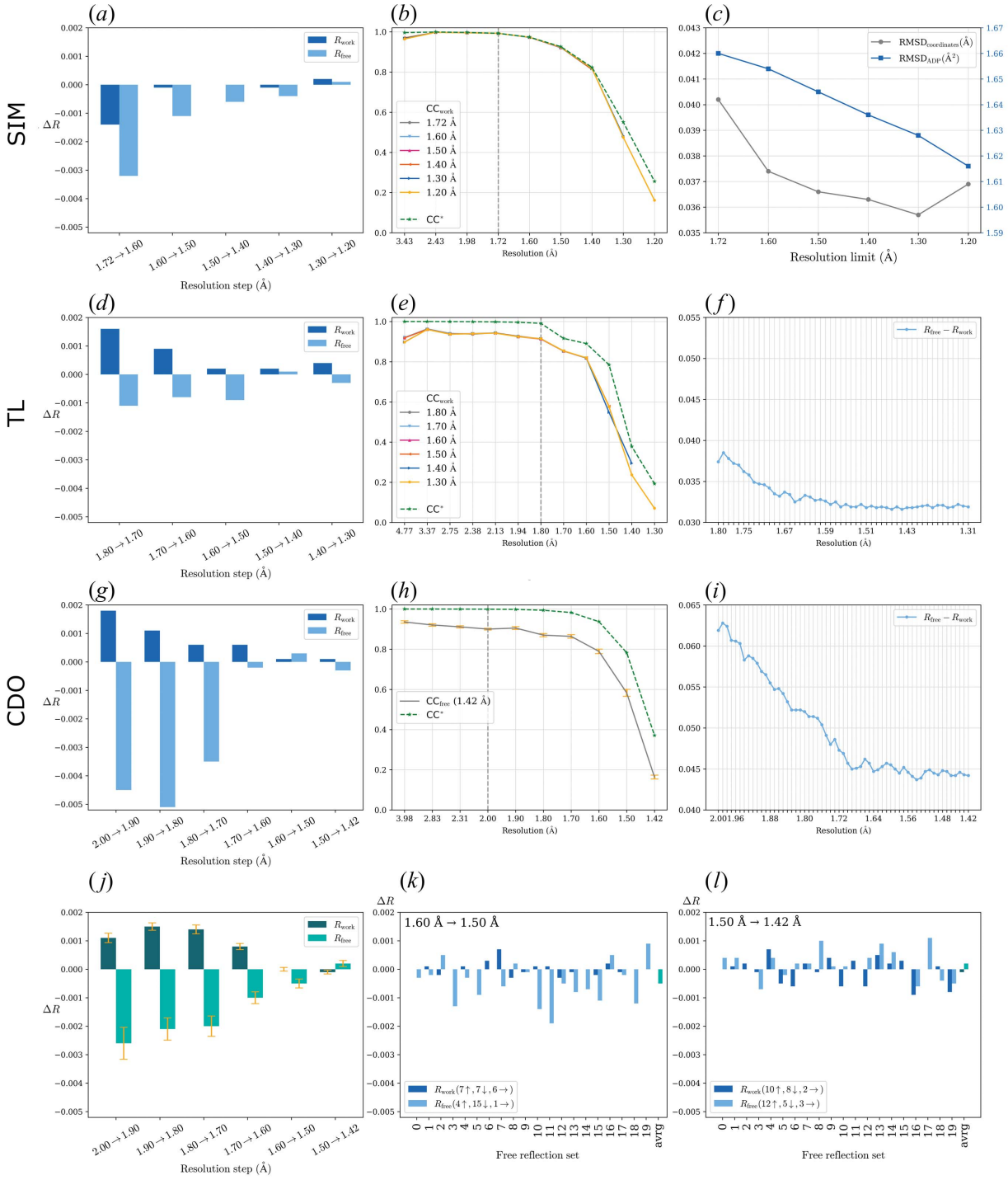


Figure 2
Results from paired refinement for SIM (*a*)–(*c*), TL (*d*)–(*f*) and CDO (*g*)–(*l*). Note for bar charts showing the differences in the overall *R* values: for each incremental step of resolution for $X \rightarrow Y$, the *R* values were calculated at resolution *X*. SIM: (*a*) differences in the overall *R* values; resolution shells with a width of 0.10 Å were added stepwise. R_{free} decreases up to 1.30 Å. (*b*) Comparison of CC* and CC_{work} of refined models. (*c*) Both RMSDs of the coordinates and the ADPs ($RMSD_{coordinates}$ and $RMSD_{ADP}$) have a decreasing trend up to 1.3 Å resolution. TL: (*d*) differences in the overall *R* values; resolution shells with a width of 0.10 Å were added stepwise. (*e*) Comparison of CC* and CC_{work} of the refined models. (*f*) R_{gap} calculated using data up to 1.80 Å depending on the high-resolution cutoff; resolution shells with a width of 0.01 Å were added stepwise (a different *PAIREF* run, see the supporting information). CDO: (*g*) differences in the overall *R* values; resolution shells with a width of 0.10 Å were added stepwise. (*h*) Comparison of CC* and CC_{free} of the model refined at 1.42 Å, averaged over all of the 20 free sets. The standard error of the mean is shown in orange. (*i*) R_{gap} calculated using data up to 2.00 Å depending on the high-resolution cutoff; resolution shells with a width of 0.01 Å were added stepwise (a different *PAIREF* run, see the supporting information). (*j*) Differences in the overall *R* values averaged over all 20 free sets. The standard error of the mean is shown in orange. (*k*) and (*l*) Differences in the overall *R* values relating to all 20 free sets, refinements at 1.50 and 1.42 Å, respectively. The numbers with arrows in the legends indicate how many rises and falls were observed while using individual free reflection sets.

2012). Although the conservative criterion for R_{meas} suggests setting the high-resolution diffraction limit to 1.80 Å, having $\langle I/\sigma(I)\rangle$ higher than 2 suggests setting the limit to 1.60 Å, but paired refinement proved that data are useful up to 1.42 Å. All refinement was previously performed using *phenix.refine* (Afonine *et al.*, 2012).

Here, we tried to reproduce the previous results in *PAIREF* which uses *REFMAC*5 as a structure refinement program. We have reprocessed the original images with *XDS*. The input structure model was prepared according to the following protocol: the protein atomic positions of the unliganded CDO structure (PDB entry 2b5h; Simmons *et al.*, 2006) were perturbed by an average of 0.25 Å with *phenix.pdbtools*; the ligand (cysteine persulfenate) was built manually with *Coot*. Subsequently, the model was refined with *REFMAC*5 at 2.00 Å resolution, solvent was added automatically using *ARP*/*wARP* (Lamzin & Wilson, 1993) followed by a manual check of the ligand and solvent and restrained refinement with *REFMAC*5. This model was later used as the input file for *PAIREF* to analyze the high-resolution shells with a width of 0.10 Å. Unlike the protocol published previously, solvent molecules were not automatically updated during paired refinement.

The differences of overall R values [Fig. 2(*g*)] indicate that the high-resolution diffraction limit may be set to 1.60 Å using our combination of software and free reflection set. However, the selection of free reflections may have an impact on the results and conclusions from paired refinement; therefore, we ran the second procedure of 20-fold cross-validation across all free reflection sets, as described in Section 2.1. The differences of overall R_{free} averaged over the free sets are negative up to 1.50 Å resolution [Fig. 2(*j*)]. CC* remains higher than CC_{work} in the whole resolution range for all the refined models. Moreover, the trend of R_{gap} [Fig. 2(*i*)] shows a moderate decrease for higher resolution going up to 1.42 Å when shells with a width of 0.01 Å were analyzed in the third run of paired refinement using the original free flag 0. To conclude, our calculations indicate that the data improve the model up to 1.50 Å resolution. This suggestion originates from the complete cross-validation protocol which should always be considered when deciding on the high-resolution cutoff.

3.4. Endothiapepsin in complex with fragment B53

In the cases reported above, the improvement of structure models using paired refinement was shown on statistical criteria. However, the increase in information gained from the data may also be shown by the interpretability of electron-density maps. Such enhancement was already reported for the crystal structure of the prokaryotic sodium channel pore (improvement from 4.0 to 3.5 Å resolution) and on the crystal structure of the YfbU protein from *E. coli* (improvement from 3.1 to 2.5 Å resolution) (Karplus & Diederichs, 2015). To demonstrate this effect using *PAIREF*, we reprocessed the diffraction data from the crystal structure of endothiapepsin (EP) from *C. parasitica* in complex with fragment B53 (PDB entry 4y4g; Huschmann *et al.*, 2016) using *XDS*. The data set originates from a fragment screening project; fragment B53 has a partial occupancy.

The data were originally processed up to 1.44 Å resolution with an $\langle I/\sigma(I)\rangle$ value of 2 in the highest resolution shell (1.52–1.44 Å). Here, we tried to simulate the regular workflow of model building and structure refinement. We removed all solvent molecules including ligands from the deposited model. The atomic coordinates were perturbed as done previously, the ADPs were manually set to their mean value of 16 Å^2. Subsequently, 15 cycles of restrained refinement using anisotropic ADPs were performed with *REFMAC*5. These procedures were later followed by *PAIREF* calculations up to a resolution of 1.05 Å. According to our results, the optimal high-resolution limit was set to 1.20 Å [Fig. 3(*a*)] since positive R_{free} differences are observed for the higher resolution shells.

Inclusion of more intensities in the working data set considerably improved the quality of the omit map belonging to the partially occupied ligand [Fig. 3(*c*)]. In general, we expect that the greatest improvement in interpretability will occur for weak density features because the noise level of the map decreases due to improved phases resulting from a more accurate model. This will not significantly influence the observation of atoms with strong density. However, for a feature in the electron-density map that is close to the lower contour levels used in interpreting the map, having a bit less noise will have a higher impact on the reliability and interpretability of the electron-density map. In our case, this effect was observed in the stage of ligand and solvent building, which may be valuable especially in difficult cases and with low-occupied ligands.

3.5. Interferon gamma

All the above-mentioned cases are high-resolution crystal structures. The crystal structure of interferon gamma from *P. olivaceus* (POLI) was previously determined at a medium resolution of 2.3 Å (Zahradník *et al.*, 2018). Moreover, the data exhibited severe anisotropy. Resolution limits were estimated in the range from 2.26 to 2.71 Å, according to the criterion of $\langle I/\sigma(I)\rangle$ being higher than 1.5 in the highest resolution shell (Evans & Murshudov, 2013). The data were reprocessed in *XDS* up to 1.9 Å resolution. The deposited structure (PDB entry 6f1e; Zahradník *et al.*, 2018) was refined using all of the reflections in the final refinement step. However, we used the last model refined using work reflections only in our paired refinement.

Several parameters were used to evaluate the high-resolution cutoff. Monitoring of R_{free} differences suggests a high-resolution cutoff at 2.0 Å [see Fig. 3(*d*)]. The value of R_{work} of the model refined at 1.9 Å calculated against the data in the highest resolution shell (2.0–1.9 Å) is high: 0.43 [Fig. 3(*f*)], *i.e.* it exceeds the R value of a perfect model refined against random data (see Section 2.2). We suggest omitting the highest resolution shell in further refinement and cutting the data at 2.0 Å resolution. Poor CC* values in the high resolution are probably caused by the anisotropy of the diffraction data which affects the correlation between reflections. These

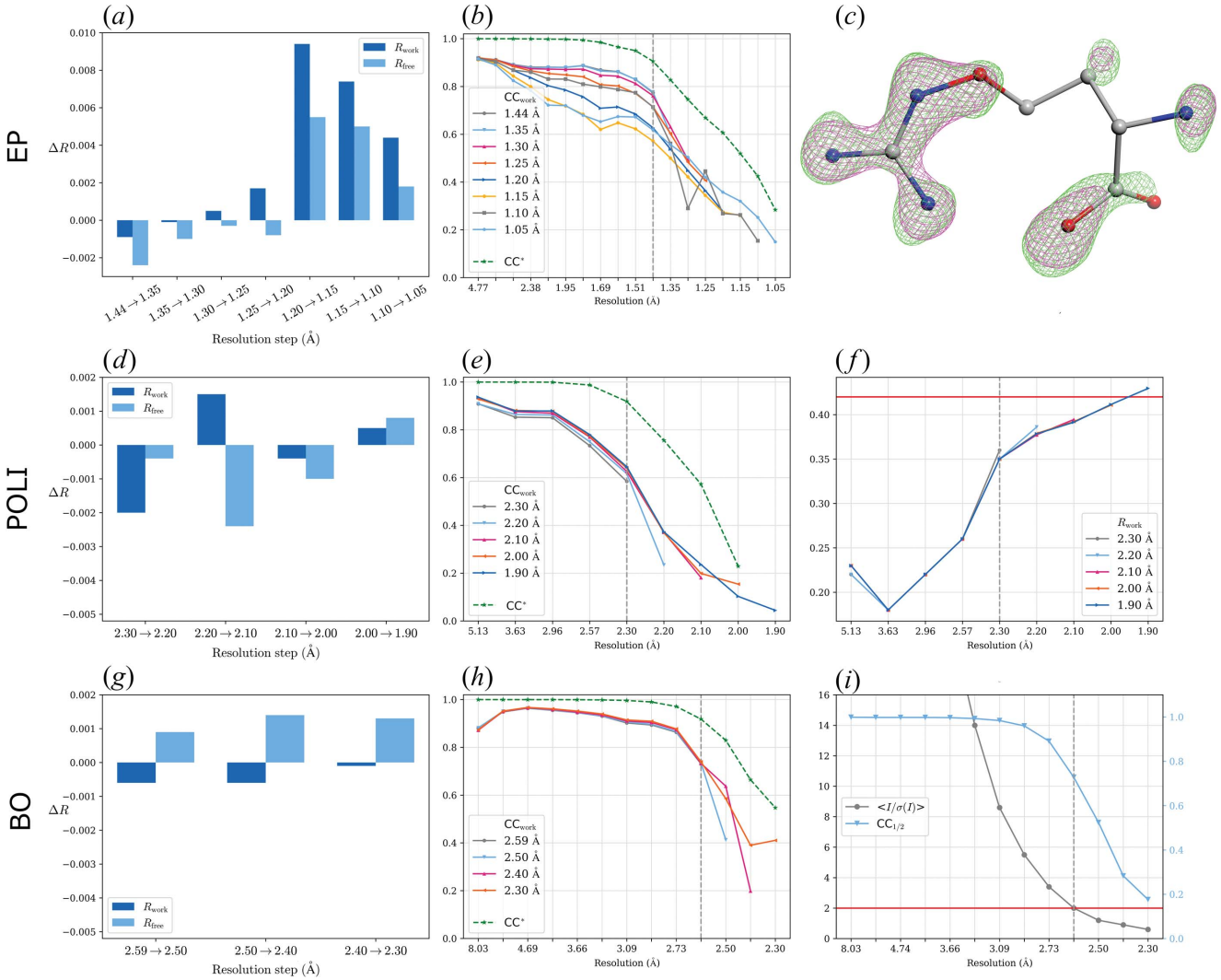


Figure 3
Results from paired refinement for EP (*a*)–(*c*), POLI (*d*)–(*f*) and BO (*g*)–(*i*). Note for bar charts showing the differences in the overall *R* values: for each incremental step of resolution for $X \rightarrow Y$, the *R* values were calculated at resolution *X*. EP: (*a*) differences in the overall *R* values; resolution shells with a width of 0.05 Å were added stepwise. A systematic decrease in R_{free} was observed up to 1.20 Å. (*b*) CC* remains higher than CC$_{work}$ in the whole resolution range for all the refined models. (*c*) Improvement in electron-density quality of the partially occupied fragment B53. Omit maps after refinement up to 1.44 (magenta) and 1.20 Å (green) are contoured at a level of 0.56 e Å^{-3}. Atomic positions of the fragment molecule originate from PDB entry 4y4g (Huschmann *et al.*, 2016). The graphic was rendered in *CCP4mg* (McNicholas *et al.*, 2011). POLI: (*d*) differences in the overall *R* values; resolution shells with a width of 0.10 Å were added stepwise. (*e*) Comparison of CC* and CC$_{work}$ of refined models. (*f*) R_{work} of refined models. The level R_{work} = 0.42 is shown as a red line. BO: (*g*) differences in the overall *R* values; resolution shells with a width of 0.10 Å were added stepwise. (*h*) Comparison of CC* and CC$_{work}$ of refined models. (*i*) $\langle I/\sigma(I)\rangle$ and CC$_{1/2}$ of the diffraction data depending on resolution; the level $\langle I/\sigma(I)\rangle$ = 2 is shown as a red line.

results show that the decision on diffraction data resolution should not be based only on a single/certain value of data quality indicator, but on a more comprehensive evaluation of the available data.

3.6. Bilirubin oxidase

The choice of the structure refinement program and parameters of refinement are the most decisive tools in paired refinement. *PAIREF* supports broad modification of structure refinement protocols using a command file for *REFMAC*5, including modification of ligand libraries. To demonstrate this functionality, we have analyzed the crystal structure of bilirubin oxidase in complex with ferricyanide (BO) (PDB entry 6i3j). The structure was previously refined at 2.59 Å resolution with $\langle I/\sigma(I)\rangle$ equal to 2 in the highest resolution shell (Koval' *et al.*, 2019) as shown in Fig. 3(*i*).

We have reprocessed the diffraction data up to a resolution of 2.3 Å with *XDS*. The last model originally refined using working reflections only was used as an input file for paired refinement. The library definitions for hexacyanoferrate, weighting matrix and several external harmonic restraints were supplied to the refinement protocol (see the supporting information). In this case, no improvement in resolution can be expected according to *PAIREF*. Although the values of CC* are higher than CC$_{work}$ and CC$_{free}$ in the whole resolution

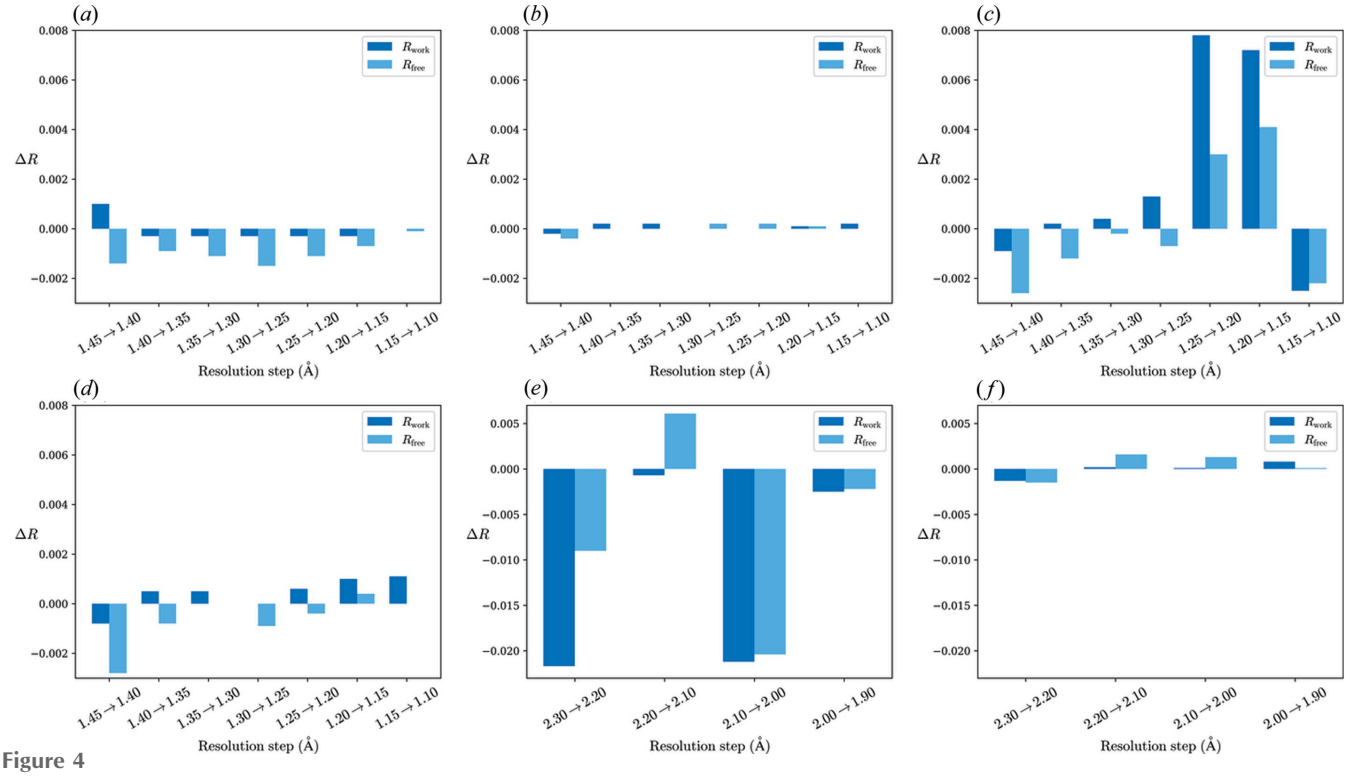


Figure 4
Paired refinement results for models from different building stages: EP (*a*)–(*d*) and POLI (*e*)–(*f*) data sets. For each incremental step of resolution for $X \rightarrow Y$, the R values were calculated at resolution X. EP: resolution shells with a width of 0.05 Å. (*a*) Model after molecular replacement using a penicillopepsin structure. (*b*) Protein model as built by *ARP*/*wARP*. (*c*) Original model of endothiapepsin without solvent molecules (PDB entry 4y4g). (*d*) Structure of endothiapepsin as deposited in the PDB. (*e*)–(*f*) POLI: resolution shells with a width of 0.10 Å. (*e*) Poly-Ala model built by *SHELXE* into the experimental map. (*f*) Complete protein model without solvent molecules.

range [Fig. 3(*h*)], an increase in R_{free} values indicates that the original high-resolution cutoff was set reasonably [Fig. 3(*g*)].

To further prove this, we ran the paired refinement protocol with 2.8 Å resolution as a starting resolution. At such low resolution, it was important to perform moderate atomic coordinate perturbation (mean shift 0.02 Å); the ADPs were set to their mean value of 35 Å^2. In this case, paired refinement suggested the data should be cut at 2.6 Å resolution, which was the original conservative cutoff (see the supporting information).

In addition, we ran the paired refinement protocol starting at 2.59 Å resolution which was not supplied with the external harmonic restraints. An apparent improvement up to 2.5 Å resolution was observed in the data quality indicators. However, refinement lacking the important restraints led to unacceptable geometry of hexacyanoferrate molecules and of several amino acid residues (away from the active site) in the output files and could not be accepted as a positive result. Analysis of the geometry of the refined model is beyond the scope of the *PAIREF* program as it is not implemented. Therefore, it remains the user's responsibility to perform such analysis. To that end, *PAIREF* provides direct links to input, output and log files from all calculation procedures.

3.7. Impact of the model quality

We performed a limited analysis of the impact of the starting model quality on results from paired refinement. We selected the EP and POLI data sets as examples of structures solved using molecular replacement and an experimental phasing method, respectively. Several models from different model building stages were used in the analysis.

3.7.1. Molecular replacement and the EP data set. We solved the structure using the molecular replacement method with *Phaser* (McCoy *et al.*, 2007). The crystal structure of penicillopepsin (54% identity, 67% similarity; PBD entry 2wea; Ding *et al.*, 1998) was used as a search model. Subsequently, the protein chain was built automatically by *ARP*/*wARP* (Langer *et al.*, 2008) at the starting resolution (1.45 Å). Altogether, we analyzed four stages of the model building: (i) model placed by molecular replacement (*i.e.* containing the penicillopepsin sequence), (ii) the protein chain built by *ARP*/*wARP*, (iii) the original model of the final structure (PDB entry 4y4g) without solvent and (iv) the final complete deposited model [Figs. 4(*a*)–4(*d*)]. We used an identical setup for all the paired refinement protocols. Initially, the coordinates were perturbed by an average of 0.25 Å and the ADPs were set to their mean value, followed by 250 refinement cycles at the starting resolution (required for refinement convergence). Then, high-resolution shells with a width of 0.05 Å were added stepwise (see the supporting information).

Surprisingly, utilization of the data in the whole resolution range (up to 1.10 Å) is suggested when using a distant protein model correctly placed in the asymmetric unit. In contrast to this, improvement only up to 1.30 Å is observed using the

model after complete protein rebuilding with *ARP/wARP*. Use of a protein model with no solvent molecules suggests the application of a high-resolution cutoff at 1.25 Å and for the most complete model at 1.20 Å.

3.7.2. Experimental phasing and the POLI data set. The crystal structure of interferon gamma from *P. olivaceus* was solved using SAD phasing. The following stages of model building were analysed: a poly-Ala model from *SHELXE* (Sheldrick, 2002), a complete protein model without solvent from *PHENIX AutoBuild* (Terwilliger *et al.*, 2008) [Figs. 4(*e*) and 4(*f*)] and the model prior to the final refinement [Fig. 3(*d*)] at the starting resolution (2.3 Å). Here we used optimized parameters of the paired refinement protocol for each specific model (see the supporting information).

The use of incomplete models in paired refinement suggested the application of a high-resolution cutoff of 2.2 Å, while the use of the most complete model a cutoff of 2.0 Å. Given both examples mentioned above, it can be stated that the model quality and completeness may play a significant role in the results from paired refinement.

4. Limitations and further development

Amongst the hundreds of trials we performed, we did not register any failure of *PAIREF* itself. However, in a few cases, the external programs may fail to report an appropriate value, which may cause the crash of the *PAIREF* run. These cases were observed mostly at unreasonable resolution, *e.g.* the third or fourth resolution shell that should have already been omitted, or during analysis of very thin shells (*e.g.* 0.01 Å).

Results of paired refinement are strongly influenced by the structure refinement protocol (and in some cases also by the specific *REFMAC5* version). In most of the cases mentioned above, a possible improvement in model accuracy owing to the use of higher-resolution data was detected using *PAIREF*. However, no improvement from the conservative cutoff was observed in the case of bilirubin oxidase.

The main focus of our further development will be the implementation of structure refinement using *phenix.refine*. Most of the procedures cannot be parallelized. Nevertheless, the parallelization of the complete cross-validation protocol is planned to significantly reduce computational time. Moreover, the inclusion of other monitoring statistics – *e.g.* $R_{\rm complete}$ (Luebben & Gruene, 2015) – in the final report is under development.

5. Discussion

In macromolecular refinement, the maximum amount of valuable data should be used to obtain the best possible structural models. Hence, evaluation of data significance should be based on novel approaches. This involves the implementation of correlation coefficients and simultaneous monitoring of trends of several statistics that are directly linked to the quality of the refined model. Paired refinement is currently generally accepted as the optimal protocol for the determination of high-resolution cutoff. The *PAIREF* program is a command-line tool that performs such an analysis and creates a compact report for users to make a self-contained decision on the data limit.

In one of the examples documented here, we first analyzed the progress of the paired refinement procedure as well as the *PAIREF* functionality on data that have been artificially generated from a known structure. This structure later served as a target to monitor the convergence of the refined models. Continuous improvement in agreement between the original structure and models from paired refinement was observed in a range where our criteria suggested acceptance of further data. Here, the RMSD calculations showed that use of the high-resolution cutoff suggested by paired refinement produces models closest to the truth. The gap between $CC_{\rm work}$ and CC* visible for all projects except SIM corresponds to the *R*-value gap discussed by Holton *et al.* (2014), and is due to deficiencies in modelling the experiment.

We also tested the program on five other real cases, some of them previously used in paired refinement. In four cases, we showed that the model could be further improved by the use of data beyond conservative cutoffs. Our program is able to successfully reproduce two particular paired refinement protocols that were published previously [TL in the work by Winter *et al.* (2018) and CDO in the work by Karplus & Diederichs (2012)] and the results obtained are in good agreement with the original ones. Slight differences could be caused by the use of a newer version of *REFMAC5* (in the case of TL), or by the utilization of other refinement software and the absence of an automatic solvent update during paired refinement (in the case of CDO).

In the case of bilirubin oxidase, an agreement in the high-resolution estimation between the conservative and paired refinement approach was observed. In all reported cases, the values of $\langle I/\sigma(I)\rangle$ and $CC_{1/2}$ are in the ranges from 0.1 to 1.7 and from 0.027 to 0.524, respectively, all in the highest accepted resolution shell. Therefore, it is clear that a resolution cutoff based purely on certain values of these statistics does not correspond to the information content in the last or next additional resolution shell, as shown in previous works (Karplus & Diederichs, 2012, 2015; Diederichs & Karplus, 2013; Evans & Murshudov, 2013; Winter *et al.*, 2018).

The addition of high-resolution reflections suggested by the paired refinement results influences the amount of experimental data used in structure refinement as well as the overall agreement of the model to the data. In addition, it produces cleaner and more detailed maps which enable further manual improvement and removal of model errors by refinement. In the case of the data set from fragment screening (EP), we demonstrated that the involvement of valid data from higher resolution shells may have a positive impact on the quality of the electron-density map. Such an effect is clearly useful for low-occupancy ligands, partially disordered regions, alternative positions or low-resolution data.

We tested the influence of model quality on the results from paired refinement. We randomly chose a distant model for molecular replacement of the structure of endothiapepsin and simulated the procedure of structure building and refinement.

We also used three models from various stages of structure determination of interferon gamma from *Paralichthys olivaceus*. In these two cases, we observed that the use of a poor starting model suggested a lower high-resolution cutoff than the use of the most complete models. This notwithstanding, the use of a (partially) incorrect model may also result in a misleading suggestion, *e.g.* inclusion of the whole resolution range. Therefore, the input structure model should be selected carefully; paired refinement is particularly sensible in the final stage of structure refinement.

PAIREF worked well for the examples described using this general protocol: (i) processing of diffraction data at (almost) the full resolution; (ii) provisional resolution cutoff according to a conservative criterion, structure solution, model building and refinement; (iii) paired refinement with sufficient model quality at a later stage of model refinement.

With the introduction of paired refinement into X-ray crystallography, the high-resolution diffraction limit has gained a new meaning, as the only criterion for the data cutoff is now the 'additional value' of the data in model refinement. Following the current trends in diffraction data evaluation, resolution cannot be directly related to a specific value of the conventional indicators of diffraction data quality.

Reflections that were added during the paired refinement protocol generally represent data with the lowest information content. Since they come from the highest resolution shells, their $\langle I/\sigma(I)\rangle$ is lower, $R_{\rm meas}$ higher and $CC_{1/2}$ lower. Nonetheless, they may represent a significant portion of the data. For most of the cases reported above, the reflections added through paired refinement account for more than 40% of all data. This of course is highly dependent on the conservative criteria that were used previously, before the paired refinement protocol was applied. Moreover, paired refinement has shown its importance for the improvement of structure models or even interpretability of electron-density maps.

Acknowledgements

We thank Andrew Karplus for comments on the manuscript, James Holton for the discussion regarding the simulated data set SIM, Jan Stránský for development consultation, Jan Wollenhaupt and Manfred S. Weiss for providing the EP data set, and Tomáš Koval' and Leona Švecová for providing the BO data set.

Funding information

This work was supported by the Ministry of Education, Youth and Sports CR – projects CAAS (grant No. CZ.02.1.01/0.0/0.0/16_019/0000778 to the to Faculty of Nuclear Sciences and Physical Engineering, Czech Technical University in Prague); ELIBIO (grant No. CZ.02.1.01/0.0/0.0/15_003/0000447 to the Institute of Biotechnology AS CR) and BIOCEV (grant No. CZ.1.05/1.1.00/02.0109 to the Institute of Biotechnology AS CR), from the ERDF fund; the Czech Science Foundation (grant No. 18-10687S to the Institute of Biotechnology AS CR); the Czech Academy of Sciences (grant No. 86652036); and by the Grant Agency of the Czech Technical University in Prague (grant No. SGS19/189/OHK4/3T/14).

References

Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C. & Zwart, P. H. (2010). *Acta Cryst.* D**66**, 213–221.
Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H. & Adams, P. D. (2012). *Acta Cryst.* D**68**, 352–367.
Behnen, J., Köster, H., Neudert, G., Craan, T., Heine, A. & Klebe, G. (2012). *ChemMedChem*. **7**, 248–261.
Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). *Nucleic Acids Res.* **28**, 235–242.
Brünger, A. T. (1993). *Acta Cryst.* D**49**, 24–36.
DeLano Scientific (2017). *The PyMOL Molecular Graphics System*, version 2.0. Schrödinger, LLC.
Diederichs, K. & Karplus, P. A. (2013). *Acta Cryst.* D**69**, 1215–1222.
Ding, J., Fraser, M. E., Meyer, J. H., Bartlett, P. A. & James, M. N. G. (1998). *J. Am. Chem. Soc.* **120**, 4610–4621.
Drenth, J. & Jeroen, M. (2010). *Principles of Protein X-ray Crystallography*, pp. 248–278. New York: Springer.
Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.
Evans, P. R. (2011). *Acta Cryst.* D**67**, 282–292.
Evans, P. R. & Murshudov, G. N. (2013). *Acta Cryst.* D**69**, 1204–1214.
Girard, E., Chantalat, L., Vicat, J. & Kahn, R. (2002). *Acta Cryst.* D**58**, 1–9.
Grosse-Kunstleve, R. W., Sauter, N. K., Moriarty, N. W. & Adams, P. D. (2002). *J. Appl. Cryst.* **35**, 126–136.
Holton, J. M., Classen, S., Frankel, K. A. & Tainer, J. A. (2014). *FEBS J.* **281**, 4046–4060.
Hunter, J. D. (2007). *Comput. Sci. Eng.* **9**, 90–95.
Huschmann, F. U., Linnik, J., Sparta, K., Ühlein, M., Wang, X., Metz, A., Schiebel, J., Heine, A., Klebe, G., Weiss, M. S. & Mueller, U. (2016). *Acta Cryst.* F**72**, 346–355.
Jiang, J.-S. & Brünger, A. T. (1994). *J. Mol. Biol.* **243**, 100–115.
Joosten, R. P., Long, F., Murshudov, G. N. & Perrakis, A. (2014). *IUCrJ*, **1**, 213–220.
Kabsch, W. (2010). *Acta Cryst.* D**66**, 125–132.
Karplus, P. A. & Diederichs, K. (2012). *Science*, **336**, 1030–1033.
Karplus, P. A. & Diederichs, K. (2015). *Curr. Opin. Struct. Biol.* **34**, 60–68.
Kleywegt, G. J. & Brünger, A. T. (1996). *Structure*, **4**, 897–904.
Koval', T., Švecová, L., Østergaard, L. H., Skalova, T., Dušková, J., Hašek, J., Kolenko, P., Fejfarová, K., Stránský, J., Trundová, M. & Dohnálek, J. (2019). *Sci. Rep.* **9**, 13700.
Lamzin, V. S. & Wilson, K. S. (1993). *Acta Cryst.* D**49**, 129–147.
Langer, G., Cohen, S. X., Lamzin, V. S. & Perrakis, A. (2008). *Nat. Protoc.* **3**, 1171–1179.
Luebben, J. & Gruene, T. (2015). *Proc. Natl Acad. Sci.* **112**, 8999–9003.
McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
McNicholas, S., Potterton, E., Wilson, K. S. & Noble, M. E. M. (2011). *Acta Cryst.* D**67**, 386–394.
Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.
Read, R. J., Oeffner, R. D. & McCoy, A. J. (2020). *Acta Cryst.* D**76**, 19–27.
Rossum, G. van (1995). *Python Tutorial*. Amsterdam: Centrum voor Wiskunde en Informatica.
Sheldrick, G. M. (2002). *Z. Kristallogr.* **217**, 644–650.

Simmons, C. R., Krishnamoorthy, K., Granett, S. L., Schuller, D. J., Dominy, J. E., Begley, T. P., Stipanuk, M. H. & Karplus, P. A. (2008). *Biochemistry*, **47**, 11390–11392.
Simmons, C. R., Liu, Q., Huang, Q., Hao, Q., Begley, T. P., Karplus, P. A. & Stipanuk, M. H. (2006). *J. Biol. Chem.* **281**, 18723–18733.
Terwilliger, T. C., Grosse-Kunstleve, R. W., Afonine, P. V., Moriarty, N. W., Zwart, P. H., Hung, L.-W., Read, R. J. & Adams, P. D. (2008). *Acta Cryst.* D**64**, 61–69.
Tickle, I. (2011). *Number of cycles in REFMAC.* https://www.mail-archive.com/ccp4bb@jiscmail.ac.uk/msg22423.html.
Vaguine, A. A., Richelle, J. & Wodak, S. J. (1999). *Acta Cryst.* D**55**, 191–205.
Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G. W., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A. & Wilson, K. S. (2011). *Acta Cryst.* D**67**, 235–242.
Winter, G. (2010). *J. Appl. Cryst.* **43**, 186–190.
Winter, G., Waterman, D. G., Parkhurst, J. M., Brewster, A. S., Gildea, R. J., Gerstel, M., Fuentes-Montero, L., Vollmar, M., Michels-Clark, T., Young, I. D., Sauter, N. K. & Evans, G. (2018). *Acta Cryst.* D**74**, 85–97.
Zahradník, J., Kolářová, L., Pařízková, H., Kolenko, P. & Schneider, B. (2018). *Fish Shellfish Immunol.* **79**, 140–152.

All the refinement runs presented in this paper were performed with *REFMAC*5 version 5.8.258.

Installation of *PAIREF*:
PAIREF depends on the *CCP*4 package. Run a following command in Unix shell (GNU/Linux, MacOS) or in *CCP4CONSOLE* (MS Windows):

```
cctbx.python -m pip install pairef --user --no-deps
```

S1 Simulated data set of lysozyme (SIM)

Perturbation of the reference structure model:

```
phenix.pdbtools set_b_iso=15 shake=0.25 1H87_single.pdb
file_name=1H87_single_shaken.pdb
```

Run with a resolution shell width of 0.10 Å (Fig. 2a-c):

```
cctbx.python -m pairef --XYZIN 1H87_single_shaken.pdb --HKLIN 1H87_dui_R.mtz -u
1H87_dui_unmerged.mtz -i 1.72 -r 1.6,1.5,1.4,1.3,1.2 --prerefinement-ncyc 13 --ncyc 6
-p 1H87_dui
```

# Shell (A)	Rwork(init)	Rwork(fin)	Rwork(diff)	Rfree(init)	Rfree(fin)	Rfree(diff)
1.72A->1.60A	0.0609	0.0592	-0.0017	0.0712	0.0681	-0.0031
1.60A->1.50A	0.0655	0.0653	-0.0002	0.0746	0.0734	-0.0012
1.50A->1.40A	0.0733	0.0733	0.0000	0.0812	0.0805	-0.0007
1.40A->1.30A	0.0848	0.0847	-0.0001	0.0938	0.0935	-0.0003
1.30A->1.20A	0.1047	0.1049	0.0002	0.1112	0.1113	0.0001

RMS coordinates differences were calculated in PyMOL using function `rms_cur()`.
RMS of ADP differences were calculated according to formula

$$RMSD_{\mathrm{ADP}} = \sqrt{\frac{\sum_{i=1}^{N}[B_{1,i} - \overline{B_1} - B_{2,i} + \overline{B_2}]^2}{N}},$$

where both structure models consist of N atoms with ADPs $B_{1,1}$, $B_{1,2}$, …, $B_{1,N}$, and $B_{2,1}$, $B_{2,2}$, …, $B_{2,N}$. Bars denote average values.

Merging statistics in resolution bins:

#shell	d_max	d_min	#obs	#uniq	mult.	%comp	<I>	<I/sI>	r_mrg	r_meas	r_pim	cc1/2	cc_ano	cc*
01	38.62	3.43	11225	1756	6.39	99.83	725.8	27.2	0.060	0.066	0.027	0.985	0.207	0.9962
02	3.43	2.43	20639	2965	6.96	99.83	264.6	18.8	0.059	0.064	0.024	0.997	0.618	0.9992
03	2.43	1.98	25070	3846	6.52	100.00	108.4	9.5	0.105	0.114	0.043	0.988	0.404	0.9970
04	1.98	1.72	30516	4336	7.04	99.93	40.8	4.0	0.235	0.254	0.094	0.971	0.156	0.9926
05	1.72	1.60	19736	3054	6.46	100.00	20.7	2.0	0.431	0.468	0.180	0.901	0.080	0.9736
06	1.60	1.50	19671	3305	5.95	99.70	13.1	1.1	0.731	0.800	0.317	0.752	0.008	0.9265
07	1.50	1.40	23094	4325	5.34	99.61	8.9	0.6	1.110	1.230	0.515	0.513	0.024	0.8235
08	1.40	1.30	16791	5345	3.14	93.44	5.3	0.3	1.890	2.233	1.153	0.179	0.021	0.5510
09	1.30	1.20	3690	2566	1.44	33.03	2.7	0.1	3.379	4.476	2.900	0.034	-0.384	0.2564

S2 Thermolysin (TL)

Run with a resolution shell width of 0.10 Å (Fig. 2d-e):

```
cctbx.python -m pairef --XYZIN 3n21_edit05_refmac1.pdb --HKLIN
AUTOMATIC_DEFAULT_free_R.mtz -u AUTOMATIC_DEFAULT_scaled_unmerged.mtz -i 1.80 -n 5 -s
0.10 -p TL_step0-10A
```

Run with a resolution shell width of 0.01 Å (Fig. 2f):

```
cctbx.python -m pairef --XYZIN 3n21_edit05_refmac1.pdb --HKLIN
AUTOMATIC_DEFAULT_free_R.mtz -u AUTOMATIC_DEFAULT_scaled_unmerged.mtz -i 1.80 -n 49 -s
0.01 -p TL_step0-01A
```

Merging statistics in resolution bins:

#shell	d_max	d_min	#obs	#uniq	mult.	%comp	<I>	<I/sI>	r_mrg	r_meas	r_pim	cc1/2	cc_ano	cc*
01	79.99	4.77	125965	1832	68.76	100.00	26.5	56.8	0.071	0.072	0.008	1.000	0.700	1.0000
02	4.77	3.37	232091	3108	74.68	99.97	28.4	56.4	0.088	0.089	0.010	1.000	0.503	1.0000
03	3.37	2.75	301747	3938	76.62	99.92	11.7	33.1	0.150	0.151	0.017	0.999	0.355	0.9997
04	2.75	2.38	342932	4589	74.73	99.59	5.0	19.5	0.255	0.257	0.029	0.997	0.231	0.9992
05	2.38	2.13	396759	5102	77.77	99.20	2.8	13.4	0.379	0.381	0.042	0.994	0.147	0.9985
06	2.13	1.94	427018	5775	73.94	98.79	1.5	8.4	0.595	0.599	0.068	0.986	0.049	0.9965
07	1.94	1.80	455635	5900	77.23	98.38	0.6	4.8	1.040	1.046	0.117	0.965	0.052	0.9911
08	1.80	1.70	403806	5435	74.30	97.96	0.3	2.8	1.741	1.753	0.200	0.723	0.010	0.9161
09	1.70	1.60	517662	6831	75.78	97.47	0.2	1.6	2.872	2.891	0.326	0.657	0.022	0.8905
10	1.60	1.50	510558	8252	61.87	91.80	0.1	0.8	4.789	4.828	0.598	0.445	0.009	0.7848
11	1.50	1.40	299987	7139	42.02	61.05	0.0	0.2	9.608	9.727	1.445	0.077	0.023	0.3781
12	1.40	1.31	84121	4154	20.25	32.19	0.0	-0.1	20.319	20.855	4.431	0.019	0.010	0.1931

S3 Cysteine dioxygenase (CDO)

Run with a resolution shell width of 0.10 Å - free flag 0 (Fig. 2g-h):

```
cctbx.python -m pairef --HKLIN CDO_R.mtz --XYZIN 2B5H_edit_refmac1.pdb -u XDS_ASCII.HKL -i 2 -r
1.9,1.8,1.7,1.6,1.5,1.42 -p CDO_step0-10A
```

Run with a resolution shell width of 0.01 Å - free flag 0 (Fig. 2i):

```
cctbx.python -m pairef --HKLIN CDO_R.mtz --XYZIN 2B5H_edit_refmac1.pdb -u XDS_ASCII.HKL -i 2 -s
0.01 -n 58 -p CDO_step0-01A
```

Run with a resolution shell width of 0.10 Å - complete cross-validation (Fig. 2j-l):

```
cctbx.python -m pairef --HKLIN CDO_R.mtz --XYZIN 2B5H_edit_refmac1.pdb -u XDS_ASCII.HKL -i 2 -r
1.9,1.8,1.7,1.6,1.5,1.42 --complete --prerefinement-ncyc 20  --prerefinement-shake-sites 0.25 --
prerefinement-reset-bfactor -p CDO_step0-10A_complete
```

Merging statistics in resolution bins:

#shell	d_max	d_min	#obs	#uniq	mult.	%comp	<I>	<I/sI>	r_mrg	r_meas	r_pim	cc1/2	cc_ano	cc*
01	41.96	3.98	30502	2013	15.15	100.00	509.9	92.4	0.029	0.030	0.008	1.000	0.444	1.0000
02	3.98	2.83	55734	3335	16.71	100.00	382.3	74.0	0.048	0.049	0.012	1.000	0.178	1.0000
03	2.83	2.31	73318	4261	17.21	100.00	152.2	40.9	0.127	0.131	0.031	0.999	0.046	0.9997
04	2.31	2.00	86730	5014	17.30	100.00	100.2	23.4	0.254	0.261	0.062	0.997	0.009	0.9992
05	2.00	1.90	40614	2360	17.21	100.00	60.5	13.0	0.412	0.425	0.101	0.993	-0.011	0.9982
06	1.90	1.80	48612	2886	16.84	100.00	31.9	6.8	0.577	0.595	0.144	0.976	0.015	0.9939
07	1.80	1.70	60936	3619	16.84	100.00	16.1	3.5	0.854	0.881	0.214	0.933	0.108	0.9825
08	1.70	1.60	65470	4548	14.40	100.00	8.9	1.9	1.251	1.297	0.340	0.783	0.144	0.9372
09	1.60	1.50	60309	5862	10.29	100.00	4.9	0.9	2.023	2.130	0.653	0.441	-0.021	0.7824
10	1.50	1.42	27549	5842	4.72	98.83	2.9	0.3	3.175	3.557	1.527	0.074	0.004	0.3712

S4 Endothiapepsin in complex with fragment B53 (EP)

Perturbation of the structure model:

```
phenix.pdbtools set_b_iso=16 shake=0.25 4Y4G_noW_noLIG.pdb
file_name=4Y4G_noW_noLIG_shaken.pdb
```

Run with a resolution shell width of 0.05 Å (Fig. 3a-c):

```
cctbx.python -m pairef --XYZIN 4Y4G_noW_noLIG_shaken.pdb --HKLIN EP_R.mtz -u
XDS_ASCII.HKL -i 1.44 -r 1.35,1.30,1.25,1.20,1.15,1.10,1.05 --prerefinement-ncyc 15 -p
EP
```

Merging statistics in resolution bins:

#shell	d_max	d_min	#obs	#uniq	mult.	%comp	<I>	<I/sI>	r_mrg	r_meas	r_pim	cc1/2	cc_ano	cc*
01	42.68	4.77	6511	1646	3.96	99.58	198.1	41.3	0.024	0.027	0.013	0.999	-0.001	0.9997
02	4.77	3.37	10618	2984	3.56	99.60	269.3	36.7	0.025	0.030	0.015	0.999	-0.098	0.9997
03	3.37	2.75	14924	3814	3.91	99.17	112.3	26.3	0.040	0.046	0.023	0.998	-0.031	0.9995
04	2.75	2.38	17054	4492	3.80	98.99	55.7	17.9	0.059	0.068	0.034	0.996	0.016	0.9990
05	2.38	2.14	17117	4812	3.56	98.61	41.7	13.8	0.074	0.087	0.045	0.993	0.031	0.9982
06	2.14	1.95	21801	5624	3.88	98.36	29.6	11.5	0.094	0.110	0.055	0.990	0.013	0.9975
07	1.95	1.80	24698	6237	3.96	98.10	15.4	7.3	0.156	0.180	0.089	0.976	0.013	0.9939
08	1.80	1.69	23067	6065	3.80	97.78	8.6	4.5	0.245	0.285	0.144	0.941	0.044	0.9847
09	1.69	1.59	25972	7046	3.69	97.40	5.5	3.0	0.361	0.423	0.217	0.871	0.043	0.9649
10	1.59	1.51	27217	6983	3.90	96.78	3.9	2.3	0.490	0.568	0.284	0.822	0.041	0.9499
11	1.51	1.44	29343	7460	3.93	96.32	2.8	1.7	0.672	0.777	0.385	0.694	0.035	0.9052
12	1.44	1.35	44398	11896	3.73	96.04	1.9	1.1	0.967	1.128	0.573	0.521	0.020	0.8277
13	1.35	1.30	30329	8072	3.76	95.49	1.3	0.7	1.359	1.585	0.806	0.385	0.032	0.7456
14	1.30	1.25	36280	9338	3.89	95.05	1.0	0.6	1.728	2.003	1.002	0.288	0.032	0.6687
15	1.25	1.20	42646	10944	3.90	94.61	0.8	0.5	2.158	2.500	1.247	0.225	0.038	0.6061
16	1.20	1.15	47423	12864	3.69	94.05	0.6	0.3	2.764	3.235	1.656	0.156	0.038	0.5195
17	1.15	1.10	58651	15099	3.88	93.26	0.5	0.2	3.569	4.138	2.070	0.099	0.037	0.4245
18	1.10	1.05	54538	15315	3.56	79.14	0.2	0.1	5.979	6.987	3.564	0.042	0.025	0.2839

S5 Interferon gamma (POLI)

Run with a resolution shell width of 0.1 Å (Fig. 3d-f):

```
cctbx.python -m pairef --XYZIN poli67_edit12_refmac1.pdb --HKLIN poli67_R.mtz -u
XDS_ASCII.HKL -i 2.3 -s 0.1 -n 4 --ncyc 10 --prerefinement-ncyc 10 -w 0.06 -p poli67
```

Merging statistics in resolution bins:

#shell	d_max	d_min	#obs	#uniq	mult.	%comp	<I>	<I/sI>	r_mrg	r_meas	r_pim	cc1/2	cc_ano	cc*
01	47.32	5.13	22369	1949	11.48	99.44	494.0	38.8	0.042	0.044	0.013	0.999	-0.214	0.9997
02	5.13	3.63	42232	3343	12.63	99.79	203.3	31.7	0.064	0.067	0.019	0.999	-0.163	0.9997
03	3.63	2.97	57138	4246	13.46	99.95	56.9	14.3	0.151	0.157	0.042	0.997	-0.166	0.9992
04	2.97	2.57	65007	5014	12.97	99.70	11.3	4.0	0.622	0.648	0.178	0.951	-0.025	0.9874
05	2.57	2.30	73887	5623	13.14	99.89	3.7	1.3	1.839	1.914	0.523	0.730	-0.014	0.9187
06	2.30	2.20	37405	2818	13.27	99.93	1.6	0.6	4.462	4.641	1.263	0.400	0.033	0.7559
07	2.20	2.10	45724	3363	13.60	99.91	0.9	0.3	7.623	7.920	2.129	0.196	-0.008	0.5725
08	2.10	2.00	49772	4021	12.38	98.51	0.3	0.1	16.989	17.721	4.963	0.027	0.008	0.2293
09	2.00	1.90	35920	4046	8.88	81.18	0.1	0.0	41.435	43.993	14.417	-0.132	-0.016	N/A

S6 Bilirubin oxidase (BO)

Run with a resolution shell width of 0.1 Å (Fig. 3g-h):

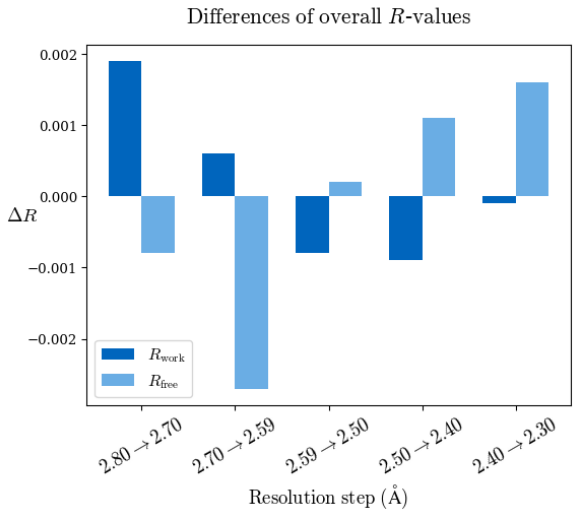
```
cctbx.python -m pairef --XYZIN BO_edit94_refmac1.pdb --HKLIN BO_R.mtz -u XDS_ASCII.HKL
--LIBIN merge_TRP-HIS_FC6.cif -c setting.com -r 2.5,2.4,2.3 -w 0.048 --prerefinement-
ncyc 10 -p BO
```

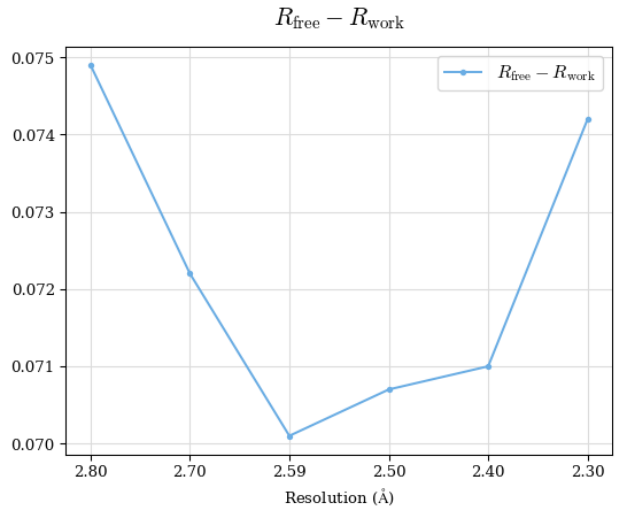
The file setting.com includes:

```
#  Hexacyanoferrate ligand definition in the working file.
# Residue numbers correspond to A613 to A616 and B611 to B615 in structure with PDB ID 6I3J, i.e. before
# renumbering in deposition
external harmonic residues from 801 A to 804 A sigma 0.03
external harmonic residues from 801 B to 805 B sigma 0.03
# protein residues
external harmonic residues from 139 B to 139 B sigma 0.01
external harmonic residues from 274 B to 274 B sigma 0.01
exte dist first chain A resi 463 atom CD second chain A resi 463 atom OE1 value 1.25 sigma 0.02
exte dist first chain B resi 492 atom CD second chain B resi 492 atom OE1 value 1.25 sigma 0.02
exte dist first chain A resi 503 atom CD second chain A resi 503 atom OE1 value 1.25 sigma 0.02
exte dist first chain B resi 143 atom CD second chain B resi 143 atom OE2 value 1.25 sigma 0.02
```

Run with a resolution shell width of 0.1 Å (2.8 Å resolution as a starting resolution):

```
cctbx.python -m pairef --XYZIN BO_edit94_refmac1.pdb --HKLIN BO_R.mtz -u XDS_ASCII.HKL
--LIBIN merge_TRP-HIS_FC6.cif -c setting.com -i 2.8 -r 2.7,2.6,2.5,2.4,2.3 -w 0.048 --
prerefinement-ncyc 10 --prerefinement-shake-sites 0.02 --prerefinement-reset-bfactor -
p BO_from2-80A
```





Merging statistics in resolution bins:

#shell	d_max	d_min	#obs	#uniq	mult.	%comp	<I>	<I/sI>	r_mrg	r_meas	r_pim	cc1/2	cc_ano	cc*
01	47.35	8.03	13105	1720	7.62	99.59	384.0	40.9	0.038	0.040	0.014	0.999	0.214	0.9997
02	8.03	5.73	23998	2921	8.22	100.00	193.6	28.5	0.063	0.067	0.023	0.998	0.092	0.9995
03	5.73	4.74	28851	3458	8.34	100.00	242.6	29.1	0.061	0.065	0.022	0.998	0.034	0.9995
04	4.74	4.10	36579	4349	8.41	100.00	258.3	27.6	0.065	0.070	0.024	0.998	0.016	0.9995
05	4.10	3.66	41914	4954	8.46	100.00	193.4	21.0	0.090	0.096	0.033	0.997	-0.002	0.9992
06	3.66	3.34	45996	5418	8.49	99.98	121.4	14.0	0.145	0.154	0.052	0.993	0.019	0.9982
07	3.34	3.09	50156	5898	8.50	100.00	73.1	8.6	0.247	0.263	0.089	0.984	0.017	0.9960
08	3.09	2.89	53641	6305	8.51	100.00	44.5	5.5	0.396	0.421	0.143	0.960	0.011	0.9897
09	2.89	2.73	54665	6426	8.51	100.00	27.6	3.4	0.650	0.691	0.235	0.891	0.016	0.9708
10	2.73	2.59	50143	7020	7.14	100.00	18.1	2.0	0.942	1.017	0.377	0.728	-0.006	0.9179
11	2.59	2.50	27420	5353	5.12	99.79	13.5	1.2	1.198	1.338	0.584	0.524	0.025	0.8293
12	2.50	2.40	27592	6777	4.07	97.54	11.2	0.9	1.396	1.613	0.785	0.283	0.008	0.6642
13	2.40	2.30	24246	7300	3.32	89.13	8.7	0.6	1.795	2.145	1.140	0.176	0.007	0.5471

S7 Impact of the model quality

S7.1 EP data set

Run with model after molecular replacement using a penicillopepsin structure (Fig 4a):

```
cctbx.python -m pairef --XYZIN EP_afterMR_2WEA.pdb --HKLIN EP_R.mtz -u XDS_ASCII.HKL -
i 1.45 -n 7 -s 0.05 --ncyc 10 --prerefinement-ncyc 250 --prerefinement-reset-bfactor -
-prerefinement-shake-sites 0.25 -p EP_afterMR_2WEA
```

Run with protein model as built by *ARP*/*wARP* (Fig 4b):

```
cctbx.python -m pairef --XYZIN EP_protein_chain_ARPwARP.pdb --HKLIN EP_R.mtz -u
XDS_ASCII.HKL -i 1.45 -n 7 -s 0.05 --ncyc 10 --prerefinement-ncyc 250 --prerefinement-
reset-bfactor --prerefinement-shake-sites 0.25 -p EP_protein_chain_ARPwARP
```

Run with original model of endothiapepsin without solvent molecules (Fig 4c):

```
cctbx.python -m pairef --XYZIN 4Y4G_noW_noLIG.pdb --HKLIN EP_R.mtz -u XDS_ASCII.HKL -i
1.45 -n 7 -s 0.05 --ncyc 10 --prerefinement-ncyc 250 --prerefinement-reset-bfactor --
prerefinement-shake-sites 0.25 -p EP_noW_noLIG
```

Run with deposited structure of endothiapepsin - PDB id 4Y4G (Fig 4d):

```
cctbx.python -m pairef --XYZIN 4Y4G.pdb --HKLIN EP_R.mtz -u XDS_ASCII.HKL -i 1.45 -n 7
-s 0.05 --ncyc 10 --prerefinement-ncyc 250 --prerefinement-reset-bfactor --
prerefinement-shake-sites 0.25 -p EP_4Y4G
```

S7.2 POLI data set

Run with poly-Ala model as built by *SHELXE* (Fig 4e):

```
Poly-Ala
cctbx.python -m pairef --XYZIN poli67-poliAla.pdb --HKLIN poli67_R.mtz -u
XDS_ASCII.HKL -i 2.3 -s 0.1 -n 4 --ncyc 10 --prerefinement-ncyc 55 --ncyc 60 -p
poli67-poliAla
```

Run with protein model without solvent molecules (Fig 4f):

```
No waters
cctbx.python -m pairef --XYZIN poli67-noW.pdb --HKLIN poli67_R.mtz -u XDS_ASCII.HKL -i
2.3 -s 0.1 -n 4 --ncyc 10 --prerefinement-ncyc 10 --ncyc 10 -w 0.06 -p poli67-nowaters
```

2.2 Journal article B)
PAIREF: paired refinement also for *Phenix* users

Firstly, the program *PAIREF* supported *REFMAC*5 for crystal structure refinement as was described in the previous work (Malý *et al.*, 2020). In this short article, we document the implementation of the support of the *Phenix.refine* refinement engine. Furthermore, we report a newly developed feature: graphical user interface for *PAIREF*.

The application of this feature is demonstrated using the diffraction data for thermolysin which were also included in the previous publication (Malý *et al.*, 2020). We discuss differences in results depending on the choice of a refinement engine with a particular focus on the complete cross-validation procedure.

Author contribution

Analogously to the previous scientific article, the author developed the code of the program *PAIREF*, designed and carried out several paired refinement runs, interpreted data and was responsible for writing the manuscript.

STRUCTURAL BIOLOGY COMMUNICATIONS

ISSN 2053-230X

Received 23 April 2021
Accepted 11 June 2021

Edited by J. Agirre, University of York, United Kingdom

Keywords: macromolecular crystallography; *PAIREF*; *Phenix*; X-ray diffraction; paired refinement; high-resolution limit.

PAIREF: paired refinement also for *Phenix* users

Martin Malý,[a,b] Kay Diederichs,[c] Jan Dohnálek[b] and Petr Kolenko[a,b]*

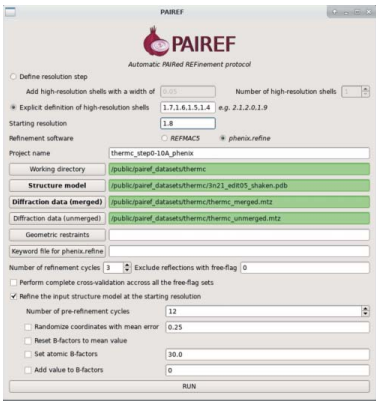
[a]Faculty of Nuclear Sciences and Physical Engineering, Czech Technical University in Prague, Břehová 7, 115 19 Prague 1, Czech Republic, [b]Institute of Biotechnology of the Czech Academy of Sciences, BIOCEV, Průmyslová 595, 252 50 Vestec, Czech Republic, and [c]University of Konstanz, Box M647, 78457 Konstanz, Germany. *Correspondence e-mail: petr.kolenko@fjfi.cvut.cz

In macromolecular crystallography, paired refinement is generally accepted to be the optimal approach for the determination of the high-resolution cutoff. The software tool *PAIREF* provides automation of the protocol and associated analysis. Support for *phenix.refine* as a refinement engine has recently been implemented in the program. This feature is presented here using previously published data for thermolysin. The results demonstrate the importance of the complete cross-validation procedure to obtain a thorough and unbiased insight into the quality of high-resolution data.

1. Introduction

During diffraction data processing, a cutoff is usually applied to reject high-resolution data that do not improve the model during structure refinement. In recent decades, a number of criteria have been used to decide on this cutoff (Karplus & Diederichs, 2015). Nowadays, paired refinement is considered to be the optimal approach for the determination of this parameter (Karplus & Diederichs, 2012; Diederichs & Karplus, 2013). In brief, a conservative cutoff has to be chosen first to secure reliable data. Data from higher resolution shells are added in a stepwise process to the model refinement and their impact on the model quality is verified. Such an approach is time-consuming and prone to errors when no automation is available. Paired refinement is particularly sensible in the later stages of structure refinement when the decision on the refinement program has already been made.

Recently, we developed an automatic tool (Malý *et al.*, 2020) for paired refinement that uses *REFMAC*5 (Murshudov *et al.*, 2011) for structure refinement. *PAIREF* provides comprehensive data analysis together with merging statistics, correlation coefficients (for example CC_{work} and CC*) and indicators of the stability of structure refinement. Here, we present a new feature of *PAIREF*: a module that performs structure refinement with *phenix.refine* (Afonine *et al.*, 2012). The algorithm does not differ from that relying on *REFMAC*5. Structure-refinement parameters for *phenix.refine* can be specified in detail through a definition file (option `--def setting.def`). Besides the current implementation of paired refinement in the *Phenix* package (Liebschner *et al.*, 2019), our tool provides additional features, for example a complete cross-validation procedure (Brünger, 1993; Jiang & Brünger, 1994). For each set of test reflections, the paired refinement protocol is run in parallel, and averaged data-quality indicators are reported. Both the standard and the complete cross-validation procedure are shown for test data. Moreover, a graphical user interface (Fig. 1) has recently been developed to simplify job execution.



2. Materials and methods

One of the previously reported cases where paired refinement using *REFMAC*5 proved to be helpful was the crystal structure of thermolysin. In the previous report (Winter *et al.*, 2018), the initial high-resolution cutoff was set to 1.8 Å and reflections were added in thin shells with a width of 0.01 Å. A decrease in R_{gap} ($R_{gap} = R_{free} - R_{work}$) was observed up to a resolution of 1.56 Å. Similarly, we carried out paired refinement while adding shells with a width of 0.10 Å, referred to as run 0 in this manuscript (Malý *et al.*, 2020). The R_{free} values systematically decreased up to a resolution of 1.5 Å, which indicates model improvement. Thus, we suggested cutting the data at this resolution level. Here, we show the results from paired refinement carried out with *PAIREF* using *phenix.refine* (*Phenix* version 1.16-3546) instead of *REFMAC*5.

We performed three distinct runs of *PAIREF* using the previously reported diffraction data for thermolysin. The diffraction images were processed with *xia*2 (Winter, 2010) employing *DIALS* (Winter *et al.*, 2018) and *AIMLESS* (Evans & Murshudov, 2013) at 1.5 Å resolution in space group $P6_122$. The input model for all runs originated from the structure of thermolysin (PDB entry 3n21; Behnen *et al.*, 2012) with water molecules. To remove model bias, the atomic coordinates were perturbed by an average of 0.25 Å and all ADPs were set to their average value with *phenix.pdbtools* (Liebschner *et al.*, 2019). Subsequently, restrained refinement was performed with *phenix.refine* at a resolution of 1.8 Å, converging sufficiently in 12 cycles. We performed three *PAIREF* runs: run 1, a standard run with the addition of high-resolution shells with a width of 0.10 Å; run 2, a fine-sliced standard run with a width of 0.01 Å; and run 3, the complete cross-validation procedure using all 20 sets of test reflections with a shell width of 0.10 Å. As an example, the command to launch run 1 (in the Unix shell or Phenix Command Prompt) is `cctbx.python -m pairef --XYZIN 3n21_edit05_shaken.pdb --HKLIN thermc_merged.mtz -u thermc_unmerged.mtz --phenix --project thermc_step0-10A_phenix --prerefinement-ncyc 12 -i 1.8 -r 1.7,1.6,1.5,1.4`; the execution of this run using the graphical interface is also shown in Fig. 1.

The results are shown in Fig. 2. The related merging statistics and details of run 0 have previously been published (Malý *et al.*, 2020).

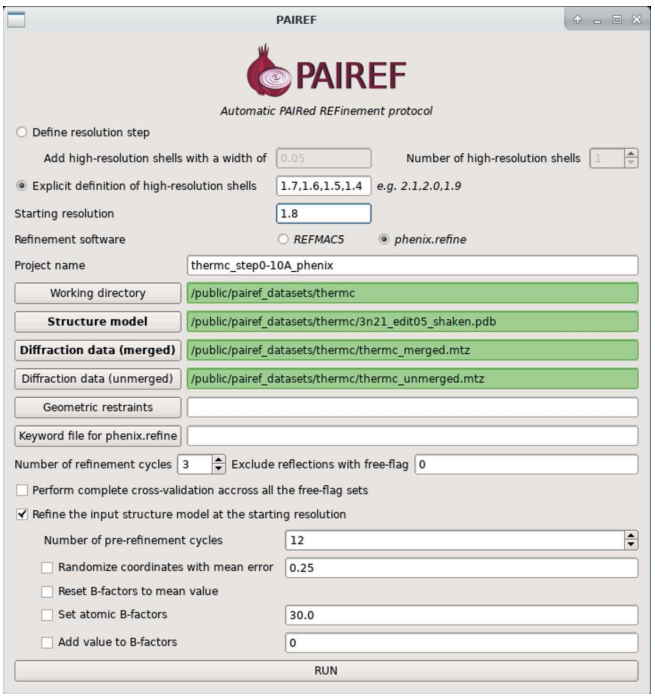


Figure 1
PAIREF graphical user interface set for the execution of run 1.

3. Results and discussion

Paired refinements using the thermolysin data demonstrate the differences that may appear using various refinement engines together with the importance of the complete cross-validation procedure (Fig. 2). Results from the individual runs, run 0 (*REFMAC*5, 0.10 Å step), run 1 (*phenix.refine*, 0.10 Å step), run 2 (*phenix.refine*, 0.01 Å step) and run 3 (*phenix.refine*, 0.10 Å step, complete cross-validation), vary in the suggestion of cutoff choice: 1.5, 1.6, 1.52 and 1.5 Å, respectively. For instance, we obtain different suggestions with *phenix.refine* and *REFMAC*5 (runs 0 and 1) while using the standard scenario, *i.e.* a 0.10 Å step and exclusion of the original set of test reflections (test flag equals 0). Moreover, the complete cross-validation procedure (run 3) suggests a higher cutoff, at 1.5 Å, than the standard run 1. As the former provides more general and meaningful insight into the quality of high-resolution data (Figs. 2*c* and 2*d*), it leads us to a final decision on the high-resolution cutoff at 1.5 Å resolution. This choice is in agreement with the fine-sliced run 2, where the R_{gap} value is minimal at 1.52 Å (Fig. 2*b*). The last resolution shell (1.6–1.5 Å) merging statistics are as follows: $I/\sigma(I) = 0.8$, $R_{p.i.m.} = 0.598$, $CC_{1/2} = 0.445$, completeness 91.8% (Malý *et al.*, 2020).

PAIREF now supports both of the most frequently used refinement programs. The refinement approach of *phenix.refine* differs from that of *REFMAC*5 in several aspects (Shabalin *et al.*, 2018), such as bulk-solvent modeling (Weichenberger *et al.*, 2015), second-derivatives approximation and separate refinement of coordinates and ADPs. Thus, certain variations in the paired refinement results could be expected. The use of *PAIREF* is not intended as a tool to decide on the choice of the refinement program, but rather as a step in structure refinement. Both refinement engines specifically treat special cases such as twinning (Campeotto *et al.*, 2018), extremes of resolution (Headd *et al.*, 2012; Kovalevskiy *et al.*, 2018), complex NCS and mixed anistropic/isotropic ADP refinement *etc*. Hence, the support for multiple refinement programs in *PAIREF* can be useful when particular data qualities need to be addressed differently (Švecová *et al.*, 2021).

The complete cross-validation procedure is recommended for thorough determination of the proper resolution cutoff. The information value of the *PAIREF* analysis could be

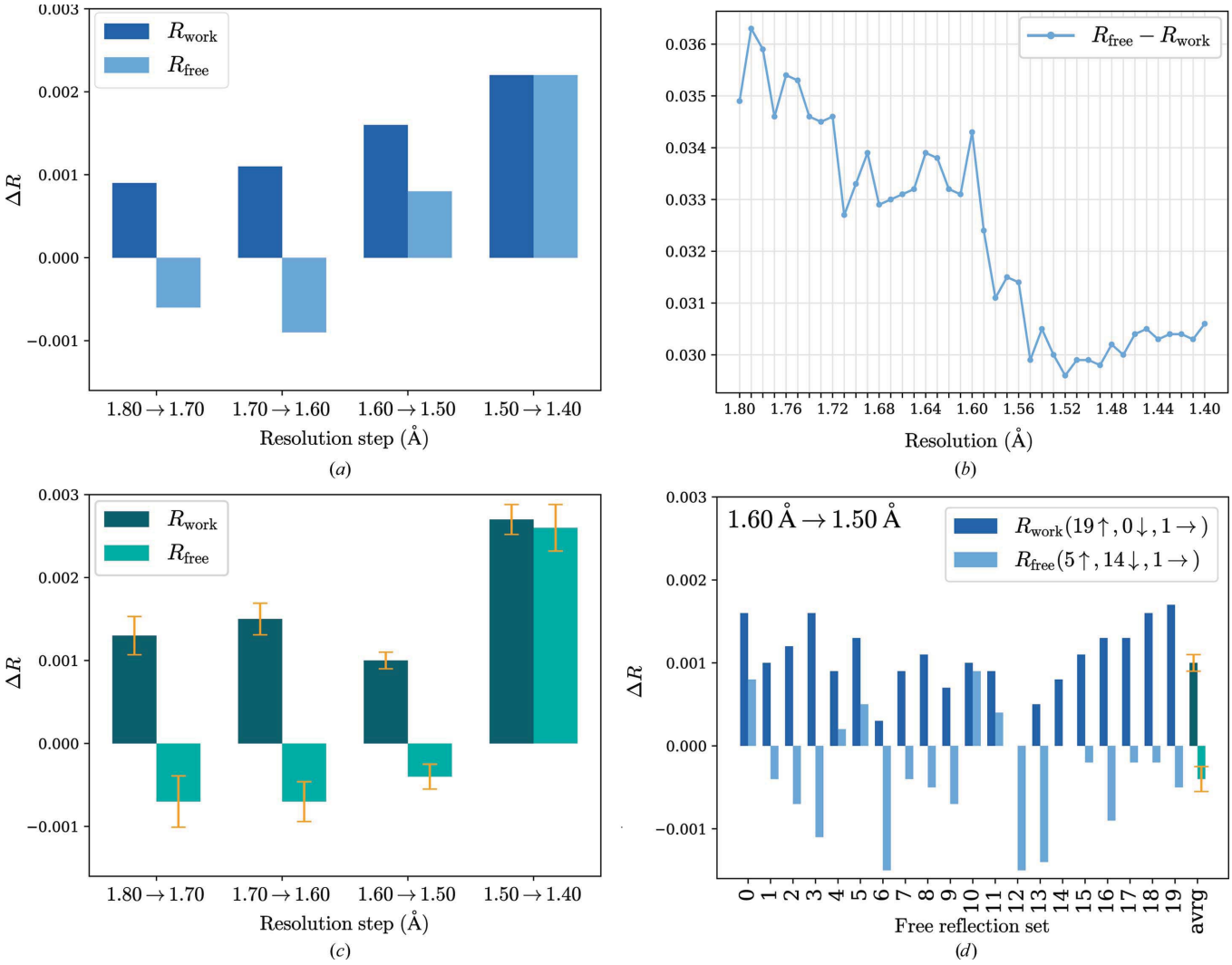


Figure 2
Report plots given by *PAIREF* from three different runs of paired refinement using *phenix.refine* with the thermolysin data. The differences in the overall R values were obtained as follows: for each incremental step of resolution increase $X \rightarrow Y$, the R values were calculated at the lower resolution X. (*a*) Run 1: differences in the overall R values; resolution shells with a width of 0.10 Å were added stepwise. R_{free} decreases up to 1.60 Å resolution, indicating model improvement. (*b*) Run 2: graph of R_{gap} calculated using data up to 1.8 Å resolution depending on the high-resolution cutoff; resolution shells with a width of 0.01 Å were added stepwise. Minimal R_{gap} is observed at 1.52 Å resolution. (*c*) Run 3: differences in the overall R values averaged over all 20 sets of test reflections. The standard error of the mean is shown in orange. (*d*) Run 3: differences in the overall R values relating to all 20 sets of test reflections for the incremental step of resolution from 1.6 to 1.5 Å. Despite the increasing R_{free} value while using the original set (test flag equals 0) and four other sets, R_{free} decreases for 14 sets. After averaging over all 20 sets, a decreasing trend is observed for this resolution shell.

further increased by the implementation of a statistic which is independent of the selection of test reflections: $R_{complete}$ (Luebben & Gruene, 2015). A detailed description of the *PAIREF* program and its algorithm, output and possibilities is provided in the primary reference (Malý *et al.*, 2020) and at the web page https://pairef.fjfi.cvut.cz.

Acknowledgements

We would like to thank Jan Stránský (Institute of Biotechnology, Czech Academy of Sciences) for helpful insights in software development and Christoph Parthier (Martin Luther University, Halle-Wittenberg) for program testing.

Funding information

The following funding is acknowledged: Ministry of Education, Youth and Sports CR (grant No. CZ.02.1.01/0.0/0.0/16_019/0000778; grant No. CZ.02.1.01/0.0/0.0/15_003/0000447; grant No. CZ.1.05/1.1.00/02.0109); Czech Science Foundation (grant No. 18-10687S); Czech Academy of Sciences (grant No. 86652036); Grant Agency of the Czech Technical University in Prague (grant No. SGS19/189/OHK4/3T/14).

References

Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H. & Adams, P. D. (2012). *Acta Cryst.* D**68**, 352–367.

Behnen, J., Köster, H., Neudert, G., Craan, T., Heine, A. & Klebe, G. (2012). *ChemMedChem*, **7**, 248–261.
Brünger, A. T. (1993). *Acta Cryst.* D**49**, 24–36.
Campeotto, I., Lebedev, A., Schreurs, A. M. M., Kroon-Batenburg, L. M. J., Lowe, E., Phillips, S. E. V., Murshudov, G. N. & Pearson, A. R. (2018). *Sci. Rep.* **8**, 14876.
Diederichs, K. & Karplus, P. A. (2013). *Acta Cryst.* D**69**, 1215–1222.
Evans, P. R. & Murshudov, G. N. (2013). *Acta Cryst.* D**69**, 1204–1214.
Headd, J. J., Echols, N., Afonine, P. V., Grosse-Kunstleve, R. W., Chen, V. B., Moriarty, N. W., Richardson, D. C., Richardson, J. S. & Adams, P. D. (2012). *Acta Cryst.* D**68**, 381–390.
Jiang, J.-S. & Brünger, A. T. (1994). *J. Mol. Biol.* **243**, 100–115.
Karplus, P. A. & Diederichs, K. (2012). *Science*, **336**, 1030–1033.
Karplus, P. A. & Diederichs, K. (2015). *Curr. Opin. Struct. Biol.* **34**, 60–68.
Kovalevskiy, O., Nicholls, R. A., Long, F., Carlon, A. & Murshudov, G. N. (2018). *Acta Cryst.* D**74**, 215–227.
Liebschner, D., Afonine, P. V., Baker, M. L., Bunkóczi, G., Chen, V. B., Croll, T. I., Hintze, B., Hung, L.-W., Jain, S., McCoy, A. J., Moriarty, N. W., Oeffner, R. D., Poon, B. K., Prisant, M. G., Read, R. J., Richardson, J. S., Richardson, D. C., Sammito, M. D., Sobolev, O. V., Stockwell, D. H., Terwilliger, T. C., Urzhumtsev, A. G., Videau, L. L., Williams, C. J. & Adams, P. D. (2019). *Acta Cryst.* D**75**, 861–877.
Luebben, J. & Gruene, T. (2015). *Proc. Natl Acad. Sci. USA*, **112**, 8999–9003.
Malý, M., Diederichs, K., Dohnálek, J. & Kolenko, P. (2020). *IUCrJ*, **7**, 681–692.
Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.
Shabalin, I. G., Porebski, P. J. & Minor, W. (2018). *Crystallogr. Rev.* **24**, 236–262.
Švecová, L., Østergaard, L. H., Skálová, T., Schnorr, K. M., Koval', T., Kolenko, P., Stránský, J., Sedlák, D., Dušková, J., Trundová, M., Hašek, J. & Dohnálek, J. (2021). *Acta Cryst.* D**77**, 755–775.
Weichenberger, C. X., Afonine, P. V., Kantardjieff, K. & Rupp, B. (2015). *Acta Cryst.* D**71**, 1023–1038.
Winter, G. (2010). *J. Appl. Cryst.* **43**, 186–190.
Winter, G., Waterman, D. G., Parkhurst, J. M., Brewster, A. S., Gildea, R. J., Gerstel, M., Fuentes-Montero, L., Vollmar, M., Michels-Clark, T., Young, I. D., Sauter, N. K. & Evans, G. (2018). *Acta Cryst.* D**74**, 85–97.

2.3 Appendix: Decision-making algorithm in *PAIREF*

In the scientific articles attached above (Malý *et al.*, 2020, 2021), an approach for the interpretation of results from paired refinement was explained. However, a user had to choose a high-resolution cutoff on his own according to a *PAIREF* comprehensive analysis in an output logfile. From the recent *PAIREF* version 1.3.3, a decision-making algorithm for automatic cutoff suggestion has been implemented. The individual high-resolution shells are progressively examined with multiple criteria. The program evaluates results with two strategies: 'strict' and 'benevolent'.

The 'strict' strategy: a shell that meets all the following criteria is accepted:

a) Overall R_{free} decreased, or R_{free} did not increase more than 0.0002 (*i.e.* remained constant) and R_{work} increased.

b) Binned R_{work} and R_{free} values in the highest resolution shell are not higher than 0.40.

c) Binned value of $CC_{1/2}$ in the highest resolution shell must be positive and CC* must be higher than CC_{work}.

d) The previous shell (*i.e.* possessing lower resolution) was accepted.

The 'benevolent' strategy is a softer modification of the 'strict' strategy. The upper limit for the binned R_{work} and R_{free} values in the highest resolution shell is 0.45, which is higher than the threshold of 0.40 in the strict criterion b). If a previous shell was not accepted due to an increase in R_{free} but all the remaining criteria are met for the currently evaluated shell, then R_{work} and R_{free} calculated at the conservative cutoff, *i.e.* starting resolution for paired refinement, are compared: If R_{free} decreased, or R_{free} did not increase more than 0.0002 and R_{work} increased, then this shell is accepted and the previous one as well.

Thus, individual shells are displayed in a table with a sign owing to the evaluation:

- accepted - the 'strict' criteria are met;
- warning - the 'benevolent' but not the 'strict' criteria are met;
- rejected - criteria are not met.

The application of the decision-making algorithm can be illustrated with the refinement of the structure model of interferon gamma from *Paralichthys olivaceus* (Zahradník *et al.*, 2018; PDB entry 6F1E). This example is documented in detail in the first article about *PAIREF* in section 3.5 and the corresponding Figure 3(d)-(f) (Malý *et al.*, 2020). Briefly, a resolution cutoff was set to 2.3 Å according to conventional criteria for statistics of diffraction data. Then, paired refinement using *PAIREF* was performed with a resolution step of 0.1 Å up to 1.9 Å.

Interpretation of the results using the decision-making algorithms is shown in Table 1. After subsequent refinements using the shells 2.3 Å → 2.2 Å and 2.2 Å → 2.1 Å, all the criteria of the 'strict' and 'benevolent' strategies were met. However, when the next shell 2.1 Å → 2.0 Å was involved in refinement, the R_{work} and R_{free} values related to this shell were 0.411 and 0.403, respectively. Thus, the 'strict' criterion b) was not complied with, whereas the 'benevolent' strategy still accepted this shell. Finally, the highest resolution shell 2.0 Å → 1.9 Å did not meet the criteria of both strategies. In summary, the high-resolution cutoff should be set at 2.1 Å according to the 'strict' strategy or at 2.0 Å using the 'benevolent' strategy. This conclusion corresponds well with the interpretation in the original article (Malý *et al.*, 2020).

Table 1: Interpretation of results of paired refinement carried out as is displayed in the *PAIREF* output logfile. Taking 'strict' or 'benevolent' criteria into account, a verdict on the resolution cutoff is 2.1 Å or 2.0 Å, respectively.

Shell	Accepted?	Reason
2.3 Å → 2.2 Å	**Yes**	Overall R_{free} decreased while using data in the shell 2.3-2.2 Å.
2.2 Å → 2.1 Å	**Yes**	Overall R_{free} decreased while using data in the shell 2.2-2.1 Å.
2.1 Å → 2.0 Å	**Warning**	R_{work} and R_{free} in high resolution are higher than 0.40 while using data in the shell 2.1-2.0 Å. Overall R_{free} decreased while using data in the shell 2.1-2.0 Å.
2.0 Å → 1.9 Å	**No**	$CC_{1/2}$ is negative or undefin. while using data in the shell 2.0-1.9 Å. R_{work} and R_{free} in high resolution are higher than 0.40 while using data in the shell 2.0-1.9 Å. Overall R_{free} increased while using data in the shell 2.0-1.9 Å.

2.4 Journal article C)
SHELIXIR: automation of experimental phasing procedures using *SHELXC/D/E*

The program *SHELIXIR*, introduced in this scientific article, is a versatile software tool that automates routines of experimental phasing. *SHELIXIR* executes numerous calculations with the *SHELXC*/*D*/*E* program package (Sheldrick, 2010, 2008; Usón and Sheldrick, 2018). It allows fast screening of several parameters, *e.g.* solvent content, alternative space groups and resolution limits (both high and low). This is particularly suitable for direct on-site phasing at synchrotron beamlines. For a simple use, graphical user interface is provided.

In this scientific article, algorithms of program routines are described. The application of *SHELIXIR* is demonstrated in several examples using different methods: single- and multi- wavelength anomalous dispersion, single isomorphous replacement and radiation-damage-induced phasing.

Author contribution

The author participated in the development of the code of the program *SHELIXIR* and contributed to the publication.

JOURNAL OF APPLIED CRYSTALLOGRAPHY

ISSN 1600-5767

Received 12 October 2020
Accepted 4 March 2021

Edited by A. Barty, DESY, Hamburg, Germany

Keywords: data processing; experimental phasing; solvents; automation; parallelization; *SHELIXIR*.

Supporting information: this article has supporting information at journals.iucr.org/j

SHELIXIR: automation of experimental phasing procedures using *SHELXC/D/E*

Petr Kolenko,[a,b]* Jan Stránský,[b] Tomáš Koval',[b] Martin Malý[a,b] and Jan Dohnálek[b]

[a]Faculty of Nuclear Sciences and Physical Engineering, Czech Technical University in Prague, Břehová 7, Prague, 115 19, Czech Republic, and [b]Institute of Biotechnology of the Czech Academy of Sciences, BIOCEV, Průmyslová 595, Vestec, 252 50, Czech Republic. *Correspondence e-mail: petr.kolenko@fjfi.cvut.cz

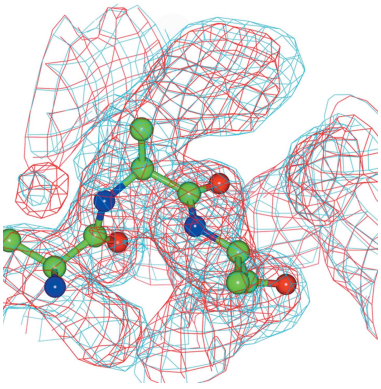
The program *SHELIXIR* represents a simple and efficient tool for routine phase-problem solution using data for experimental phasing by the single-wavelength anomalous dispersion, multiwavelength anomalous dispersion, single isomorphous replacement with anomalous scattering and radiation-damage-induced phasing methods. As indicated in its name, all calculation procedures are performed with the *SHELXC/D/E* program package. *SHELIXIR* provides screening for alternative space groups, optimal solvent content, and high- and low-resolution limits. The procedures of *SHELXE* are parallelized to minimize the computational time. The automation and parallelization of such procedures are suitable for phasing at synchrotron beamlines directly or for finding the optimal parameters for further data processing. A simple graphical interface is designed to make use easier and to increase efficiency during beam time.

1. Introduction

The phase problem is one of the major challenges in macromolecular crystallography. Despite the fact that most of the experiments performed at macromolecular beamlines are focused on the collection of high-resolution native data sets, experimental phasing is an irreplaceable method for determining novel crystal structures (Hendrickson *et al.*, 1985; Weeks *et al.*, 2003).

Experimental phasing has special technical requirements, *e.g.* for the wavelength of the primary beam or the application of heavy elements in soaking. Frequently, collection of more data sets is necessary for successful phasing. In the case of multiwavelength anomalous dispersion (MAD) data collection, radiation damage may cause a loss of information that is necessary for successful structure determination. Moreover, it is difficult to predict the phasing capacity of the data set that has been just collected. Usually, multiple data sets coming from various crystals are collected and processed after the return from the experimental time provided at synchrotrons.

Although there are many tools to predict the possibility of solving the structure [*e.g. XDS* (Kabsch, 2010) or *phenix.anomalous_signal* (Adams *et al.*, 2010)], not every data set with reasonable indicators leads to successful phasing. A new classification method using machine learning provides a high-accuracy prediction of the phasing power of data (Vollmar *et al.*, 2020). However, the best indicator of data quality is actually the ability to solve the structure and prove the solution by refinement. Given the aspects above, data collection for experimental phasing usually represents a much more time-consuming experiment than single-wavelength structure-determination approaches (*e.g.* molecular replacement), and



further development of rapid pipelines is still an important goal of software engineering. Rapid phasing with increased efficiency provides additional experimental time for focusing on native data-set collection for structure refinement and better usage of experimental time in general (Weinert *et al.*, 2015).

Many programs for experimental phasing have already been developed. Most of them are distributed within the *CCP4* (Winn *et al.*, 2011) or *Phenix* (Adams *et al.*, 2010) program packages. One of these is *SHELXC*/*D*/*E* (Sheldrick *et al.*, 2012; Sheldrick, 2010, 2008; Usón & Sheldrick, 2018), a set of programs designed as a robust tool to determine the initial phases using the single-wavelength anomalous dispersion (SAD), MAD, single isomorphous replacement (SIR), SIR with anomalous scattering (SIRAS) and radiation-damage-induced phasing (RIP) methods. While the package uses some simplifications in the calculation procedures, it is very fast and either is parallelized or can be parallelized to some extent. Furthermore, the overall requirements for RAM memory are low in comparison with other computational options. Excellent graphical user interfaces for *SHELXC*/*D*/*E* have already been developed, including *HKL2MAP* (Pape & Schneider, 2004) and *XDSgui* (https://strucbio.biologie.uni-konstanz.de/xdswiki/index.php/XDSGUI).

In experimental phasing approaches, several difficulties may occur: data incompleteness affecting correct space group determination, ambiguity of screw axes, inclusion of weak or pathologic data harming the phasing information, or incorrect estimation of the solvent content (SC) value. Many of the important parameters cannot be estimated directly from the data processing. Frequently, a number of computation trials have to be performed to find the optimum. The parameters are known after the structure is solved and refined to acceptable *R* values.

In this article, we present a new automation tool for experimental phase determination – *SHELIXIR*. The script was initially designed for rapid analysis of data at synchrotron sources because it allows a wide parallelization of *SHELXE* procedures and speeds up the process of structure determination significantly. Many synchrotron beamlines provide access to high-performance CPU clusters. Such hardware makes *SHELIXIR* suitable for fast on-site analysis of the collected data. Some synchrotron beamlines already provide pipelines attempting automatic structure solution [*e.g. ISPyB* using *FastEP* and *BigEP* (Fisher *et al.*, 2015)]; however, the scope of the attempts is limited. *SHELIXIR* provides access to all phasing methods available in *SHELXC*/*D*/*E* and searches a wider parameter space. *SHELIXIR* can also be used for teaching purposes.

2. Methods

SHELIXIR is a software tool that runs under various GNU/Linux platforms. Its dependencies are limited to Bash, *GNUplot* (http://www.gnuplot.info/) and the *SHELXC*/*D*/*E* package, which allows out-of-box application of the program. Moreover, a graphical user interface, developed in the Tcl/Tk toolkit (https://www.tcl.tk/), is provided to simplify the job execution. The program with a brief description is available online at https://kipl.fjfi.cvut.cz/shelixir.

2.1. Basic workflow

SHELIXIR works as a command-line controller of multiple *SHELXC*/*D*/*E* runs. Initially, a new directory is created and all input files are copied. In the second step, an attempt to read out the unit-cell parameters is made. If this procedure fails, it is necessary to supply the unit-cell parameters. The user must specify at least one space group. Either the space group is entered according to the user's choice (based on previous knowledge or an estimate for the given data set) or only Bravais types of lattices are entered according to the *XDS*

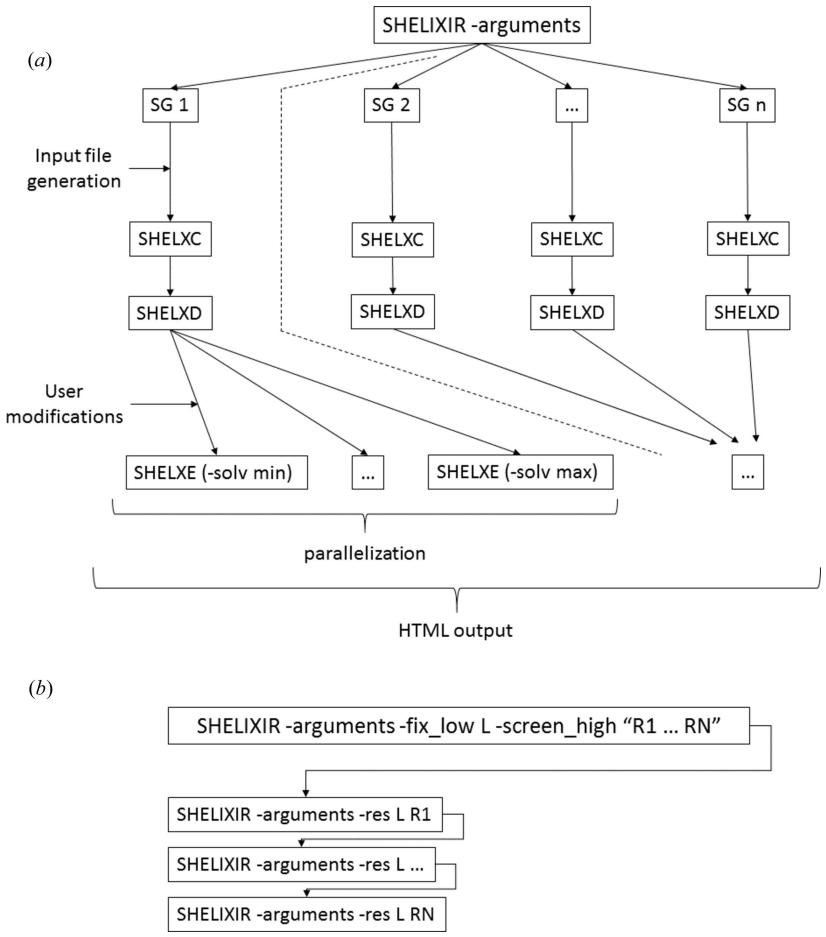


Figure 1
A schematic diagram of the *SHELIXIR* workflow. (*a*) Searching in multiple space groups (SG); 'solv min' and 'solv max' denote the minimal and maximal SC values for the *SHELXE* routine. (*b*) Screening for the optimal high-resolution limit. The limits to be screened are given as *L* for the constant low-resolution limit and *R*1 up to RN for the optimized high-resolution limit. An analogous protocol is run for the low-resolution-limit screening.

notation. The following calculations are performed for each input space group or space group belonging to the given Bravais type(s) separately.

For each space group, the *SHELXC* procedure supplied with the command-line arguments is run initially. This is followed by *SHELXD* performing the search of heavy-atom locations. When no further screening is considered, phasing and density modification are performed in *SHELXE* and the whole process starting from *SHELXC* can be repeated in the next space group.

Besides the search for the correct space group, *SHELIXIR* enables three screening procedures: search for an optimal SC value, search for an optimal high-resolution limit and search for an optimal low-resolution limit for *SHELXD* procedures. The first procedure is built into the main core of the script, whereas the latter screens are designed as a call of individual *SHELIXIR* runs with changed resolution limits. The parallelized search for an optimal SC value surpasses the optimization of the SC value that is built into other phasing software, *e.g.* the *autoSHARP* protocol (Vonrhein *et al.*, 2007). Out of all screening procedures, only the high- and low-resolution-limit screening cannot be run together in the current version. The overall workflow scheme is shown in Fig. 1.

2.2. Parallelization of procedures

The *SHELXC* procedure is usually a very fast process and requires a fraction of the time needed for all other procedures. Therefore, this procedure is not parallelized. The *SHELXD* procedure is already parallelized and uses the maximum of the computation power of the provided hardware. The *SHELXE* procedure is not parallelized internally. In *SHELIXIR*, multiple runs of the *SHELXE* procedure are run in parallel, especially for the screening of the optimal SC value, and the results are displayed after these procedures are finished. The computation time of such a subroutine can be decreased by a factor equal to the number of available CPUs.

2.3. Program output and interpretation

As a monitoring output, *SHELIXIR* produces an HTML page for each run consisting of all input parameters summarized in a table (*e.g.* a list of space groups to be tested), direct links to input and output files from *SHELXC*/*D*/*E* procedures, and a number of graphs for each space group.

In the first graph, the strength of the anomalous signal is analyzed in the 'd'' versus resolution' plot (source output file from *SHELXC*). The second and third graphs characterize the success of the *SHELXD* procedure with the 'CC_{all} versus CC_{weak}' plot and a plot of individual site occupancies. Additionally, a link to a 'PATFOM versus CC_{all}' plot is provided (PATFOM = Patterson figure of merit), but the plot is not explicitly shown on the HTML page.

The graphs and information provided by the *SHELXE* procedure differ according to the protocol. When the standard (no solvent screening) *SHELXE* run is performed, the 'contrast versus cycle' plot is used, a link to the 'connectivity versus cycle' plot is provided and the number of atoms built for both solution hands is shown. A significant discrepancy between the numbers of atoms may indicate that the correct solution was found. When the solvent-screening protocol is used, the plots of final 'contrast versus solvent content' and 'atoms built versus solvent content' are shown. The latter graph is a novel approach to the evaluation of success in experimental phasing. The graphs are generated with *GNUplot*.

The HTML page is created at the very beginning of the run and is continuously updated when partial calculations are finished. This allows the user to sort the space groups to be tested in the order of their probability and/or expectation, and to stop the calculations when the correct solution is found. Appropriate log files for each subroutine are easily accessible on the HTML page or in the directory with the results.

2.4. Preparation of the input files for test cases

The search for raw diffraction images was performed in the public repositories Zenodo (https://www.zenodo.org), SBGrid Databank (Meyer *et al.*, 2016) and the Integrated Resource for Reproducibility in Macromolecular Crystallography (Grabowski *et al.*, 2016). From a number of results, the data sets were chosen randomly. The diffraction images were processed with *XDS* in the lowest-symmetry space group (*e.g.* *P*4 for the tetragonal primitive Bravais type) to the maximal resolution and the obtained *XDS_ASCII.HKL* files were directly used as input files for *SHELIXIR*. The command-line arguments for each run are shown in the supporting information. All the results are available from http://doi.org/10.5281/zenodo.4080381.

3. Results and discussion

SHELIXIR was thoroughly tested on a number of cases. Here, we selected nine structures that vary in resolution, phasing protocol and space group. The main diffraction data and phasing characteristics are shown in Table 1.

3.1. SAD: heme PAS sensor domain of Ec DOS

A single-wavelength data set optimized for iron heavy atoms was collected for the structure determination of the heme PAS sensor domain of Ec DOS. The structure was solved in the space group $P2_12_12_1$ and refined at 1.9 Å resolution (Kurokawa *et al.*, 2004). The diffraction data were deposited in SBGrid under the reference number 137.

The data processing was performed in the space group *P*222. We ran *SHELIXIR* for all eight space groups with an orthorhombic Bravais lattice with screening for an optimal SC value. The solution was found in the correct space group $P2_12_12_1$.

Maximal connectivity, the reported value in standard procedures (Usón & Sheldrick, 2018), was observed for the SC value of 60% (model SC60). However, the maximal number of atoms built with the *SHELXE* procedure was observed for SC = 35% (model SC35). Both obtained structure models were refined with ten macrocycles in *phenix.refine*

Table 1
A summary of data processing for test structures.

The space group as deposited is shown first and *d* denotes the resolution for phasing.

Protein	PDB ID	Method	Element	*d* (Å)	Space group	*SHELXE* parameters†	Optimal/real SC (Berman *et al.*, 2000)	No. of amino acid residues built with *SHELXE*/deposited	Diffraction data DOI
Ec DOS	1v9z[a]	SAD	Fe	2.4	$P2_12_12_1$	`-h -a10 -m10 -e -z -q`	35/45	212/334	10.15785/SBGRID/137
Plectin	2odv[b]	2-w MAD	Se	2.9	$P2_122_1$	`-h -a5 -m5 -e -z -s0.5`	N/A	177/241	10.5281/zenodo.47722
Streptavidin-like protein	4jgl[c]	3-w MAD	Se	1.6	$P6_1$ ($P3_1$)	`-h -a5 -m20 -e -z`	35/49	146/186	10.18430/M34JGL
DHHW family protein	4nzk[c]	3-w MAD	Se	1.8	$P6_122$	`-h -a5 -m20 -z`	55/48	297/377	10.18430/M34NZK
Thaumatin	N/A	4-w MAD	Br	2.1	$P4_12_12$ ($P4_1$)	`-h -a5 -m5 -z`	30/56	194/207‡	Publicly available§
GerE	1fse[d]	5-w MAD	Se	3.3	$C2$	`-h -a1 -m10 -z`	40/46	296/444	N/A
N-terminal domain of UL21	4u4h[e]	SIRAS	Hg	2.0¶	$P6_322$	`-h -a5 -m10 -z`	50/48	142/213	10.15785/SBGRID/140, 10.15785/SBGRID/141
Helicase-like transcription factor	4s0n[f]	SIRAS	Se	2.6††	$P2_1$	`-h -a3 -m5 -z`	35/36	397/541	10.15785/SBGRID/129, 10.15785/SBGRID/130
Thaumatin	N/A	UV-RIP	S	1.9	$P4_12_12$ ($P4_1$)	`-h -a5 -m5 -z -q`	40/56	196/207‡	Publicly available§

References: (*a*) Kurokawa *et al.* (2004); (*b*) Sonnenberg *et al.* (2007); (*c*) JCSG (to be published); (*d*) Ducros *et al.* (2001); (*e*) Metrick *et al.* (2015); (*f*) Kile *et al.* (2015). † Parallel screening for an optimal SC using parameter `-solv 30 70 5 -nproc 11` was performed, except for the plectin case. ‡ Full sequence length. § Diffraction data and files for all tutorial projects are available at https://www.helmholtz-berlin.de/forschung/oe/np/gmx/tutorial/index_en.html. ¶ Manual, not automatic decision. †† After screening for optimal high-resolution limits using parameters `-fix_low 999 -screen_high "2.0 2.2 2.4 2.6 2.8 3.0"`.

at a resolution of 1.9 Å. The R_{work}/R_{free} values for models SC60 and SC35 were 0.43/0.47 and 0.42/0.46, respectively, indicating that the latter model produced with an unexpected SC value is better for structure building. Moreover, the quality of experimental electron-density maps should be considered. In agreement with the above-reported *R* values, the quality of experimental electron density is higher and more details are apparent for model SC35. Sections of the electron-density map are shown in Figs. 2(*e*) and 2(*f*).

The variation of results with different SC values is not counterintuitive. A high SC value (60%) allows the flattening procedure to maximize the improvements of phases and signal in the density, which is reflected in the connectivity. This is, however, at the expense of some experimental density being 'deleted' and thus not available for automated model building at a sufficient level. Therefore, the described approach could be used in optimization of the first experimental density for the automation of model building, given the used solution is consistent with the most likely one.

3.2. MAD: a fragment of the plakin domain of plectin (Cys→ Ala mutant)

The structure of a fragment of the plakin domain of plectin (Cys → Ala mutant) was solved using two-wavelength Se–Met MAD (Sonnenberg *et al.*, 2007). Although the raw data were available, we used the files from data processing using *XDS* that were deposited with Zenodo together with the diffraction images. Here, we ran the *SHELIXIR* routine without screening for the optimal SC value to demonstrate an alternative simpler approach in phasing. Out of the eight possibilities with the orthorhombic lattice, the correct solution was found in the space group $P2_122_1$ and corresponded to the original deposited structure [Protein Data Bank (PDB) ID 2odv; Sonnenberg *et al.*, 2007]. The graphs representing successful structure determination are shown in Figs. 3(*a*) and 3(*b*).

3.3. MAD: structure genomics project

Two unpublished structures determined at the Joint Center for Structural Genomics (JCSG) were selected as an example of Se-MAD phasing using three wavelengths (peak, inflection point and high-energy remote). They are the structures of hypothetical streptavidin-like protein BACEGG_01519 from *Bacteroides eggerthii* DSM 20697 and the DHHW family protein (named after the DHHW motif in the primary structure) EUBSIR_00411 from *Eubacterium siraeum* DSM 15702. In both cases, *SHELIXIR* was used to search for the phases in 21 space groups of the hexagonal Bravais type. For the hypothetical streptavidin-like protein, the successful solution was found in the correct (deposited) space group $P6_1$ and also in the lower-symmetry space group $P3_1$. Successful solution for the DHHW family protein was found only in the space group $P6_122$ as deposited in the PDB. No successful solution in a lower-symmetry space group (*e.g.* $P6_1$) was found. Both proteins were determined at high resolution and represent a routine case of structure determination.

3.4. MAD: thaumatin

The structure determination of thaumatin using bromide MAD was published within a tutorial for learning and teaching macromolecular crystallography (Faust *et al.*, 2010). The data contain the anomalous signal of bromide ions and were collected at four different wavelengths (peak, inflection point, and high- and low-energy remote).

In this case, *SHELIXIR* was successful in two space groups: the lower-symmetry space group $P4_1$ and the higher-symmetry

(correct) space group $P4_12_12$. The solution in the space group $P4_12_12$ has higher contrast than that in $P4_1$. Approximately twice as many atoms were built in the space group $P4_1$, lacking the twofold rotation axes and having a twice larger asymmetric unit. The R_{work}/R_{free} values of the two solutions are comparable. Graphs representing successful structure determination are shown in Figs. 3(*c*) and 3(*d*).

3.5. MAD: regulatory protein GerE from *Bacillus subtilis*

The crystal structure of the regulatory protein GerE from *B. subtilis* in the space group *C*2 (the only possibility in the given Bravais type monoclinic base-centred lattice) was determined using a four-wavelength selenomethionine MAD experiment together with native data up to 2.15 Å resolution. The input files for phasing are freely available in the *Phenix* program package (Adams *et al.*, 2010) and were also used for demonstration of MAD phasing using *SHELXC*/*D*/*E*. As the raw data are not freely available, we used the pre-processed files to perform the experimental phasing. The structure was successfully solved with *SHELIXIR*. To compare the computational time with the previous report (Sheldrick, 2010), the following protocol was carried out using our Intel i9-9900K CPU @ 4.6 GHz, 16 GB RAM: 100 trials to find the

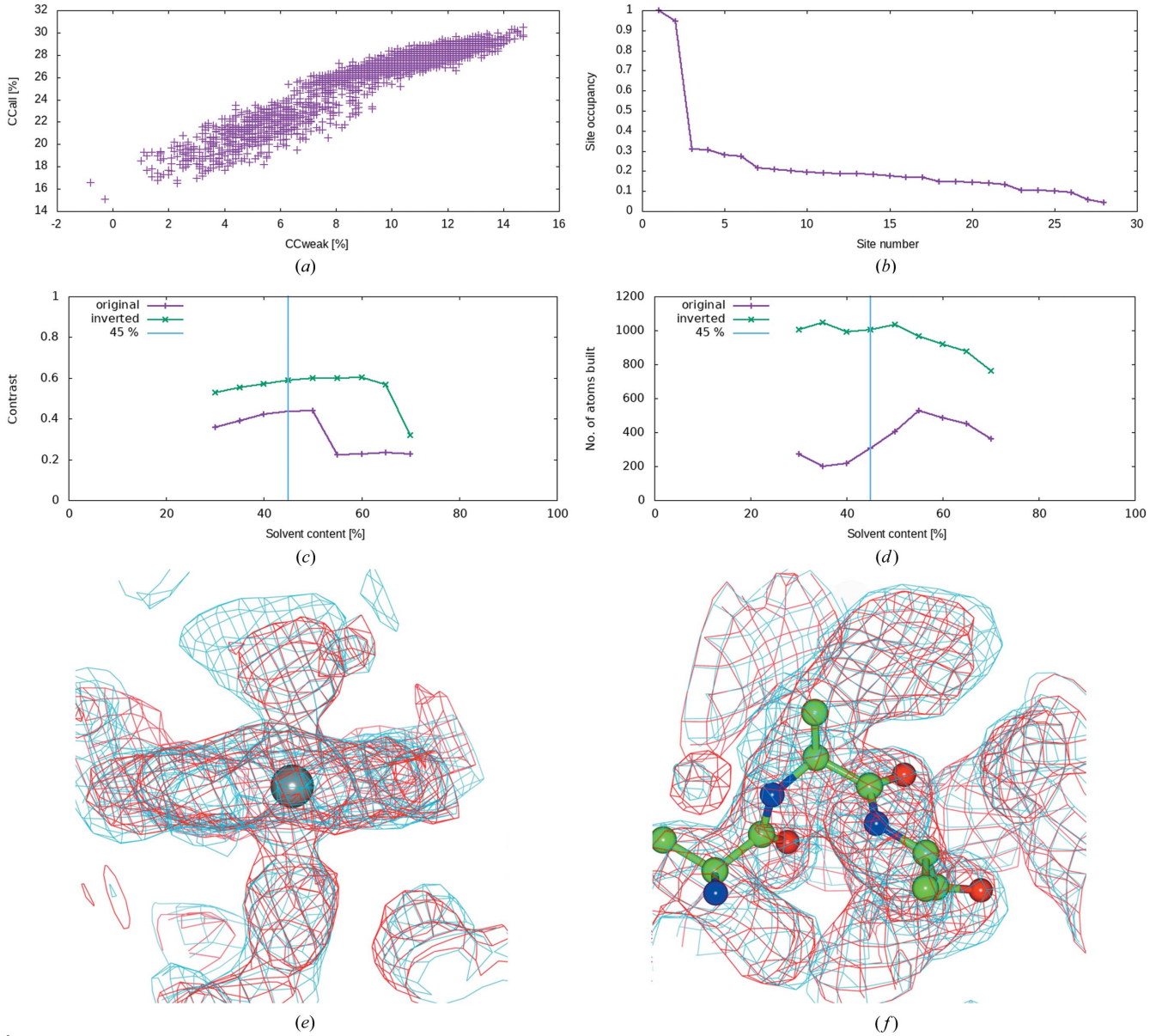


Figure 2
Successful determination of the experimental phases for the crystal structure of heme PAS sensor domain of Ec DOS. (*a*) CC_{all} versus CC_{weak} plot. (*b*) Occupancy of individual sites. (*c*) The highest contrast for the inverted hand solution is found for an SC value of 60%. (*d*) The maximum number of built atoms is observed for the SC value of 35%. Vertical lines in graphs (*c*) and (*d*) denote the real SC of 45%. Differences in the experimental electron density at a level of 1 e Å^{-3} for models SC35 (cyan) and SC60 (red) are shown for the heme region (*e*), and for the protein region (*f*). Part of the poly-Ala model is shown in stick-and-ball representation; the iron heavy atom found with the *SHELXD* routine is shown as a sphere. The figures were rendered in *CCP4mg* (McNicholas *et al.*, 2011).

substructure, one cycle of autotracing and ten cycles of density modification with screening for the optimal SC value. The computational time was under 90 s, about twice as short as in the previous report with no SC screening and older hardware.

We also tried to analyze SAD phasing capabilities using *SHELIXIR* and the 'peak' data set in a similar way. Phasing was successful again and the results were acquired within a comparable time. Here, the advantage of using *SHELIXIR*

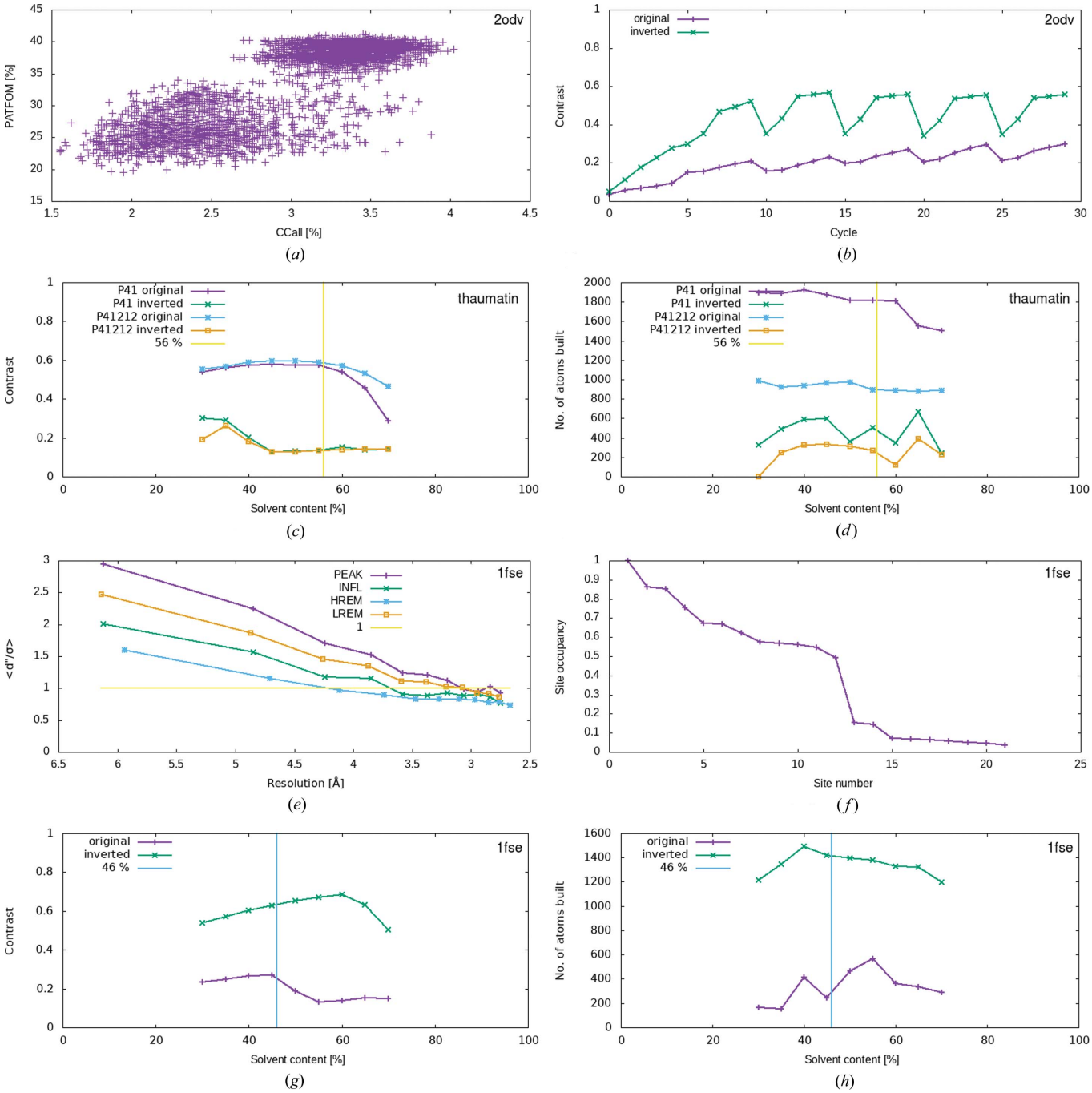


Figure 3
Characteristics of successful experimental phasing. Cys–Ala mutant of the plakin domain of plectin: (*a*) a cluster of searches distant from random searches in the 'PATFOM versus CC_{all}' plot from the *SHELXD* procedure indicates a successful search for heavy-atom substructure, and (*b*) a significant difference between the original and inverted solution in the 'contrast versus cycle' plot from the *SHELXE* procedure indicates successful model building. Thaumatin (macromolecular-crystallography tutorial, MAD): (*c*) a 'contrast versus solvent content' plot of phasing in space groups $P4_1$ and $P4_12_12$, and (*d*) the number of atoms built with the autotracing procedure in space groups $P4_1$ and $P4_12_12$. The higher number of built atoms in space group $P4_1$ is caused by the asymmetric unit being larger than that in space group $P4_12_12$. Regulatory protein GerE: (*e*) resolution dependence of mean $||F_+| - |F_-||$ divided by its standard deviation for individual data sets at various wavelengths, (*f*) occupancy of localized heavy-atom sites and (*g*), (*h*) the main characteristics of the successful main-chain tracing procedure using *SHELXE*. Vertical lines in the graphs (*c*), (*d*), (*g*) and (*h*) denote the real (as calculated for the PDB entry) SCs of individual structures.

lies in a faster identification of a successful solution, which saves further effort and experimental time. For example, with such an intense anomalous signal, the data for other wavelengths for experimental phasing would not have to be collected.

3.6. SIRAS: N-terminal domain of UL21

The structure of the N-terminal domain of UL21 from HSV-1 is a tutorial example for *autoPROC* (Vonrhein *et al.*, 2011). The structure was solved using heavy-atom derivatization with mercury ions. We processed the data up to the maximal resolution allowed by the experimental setup (1.3 Å) with *XDS* and directly used the output *XDS_ASCII.HKL* file for *SHELIXIR* calculations. Searching for the optimal SC value was a key feature that helped to solve the phase problem [see Figs. 4(*a*) and 4(*b*)]. Only a limited range of SC values led to success. We selected a model originating from calculations with the SC value adjusted to 50%, refined it with *phenix.refine* and performed automatic model building with *ARP/wARP* (Langer *et al.*, 2008), which resulted in a model with R_{work}/R_{free} values of 0.23/0.29 against the data processed at 2.0 Å resolution. We also tried to perform the same approach with models originating from the *SHELXE* procedure and with 40 or 60% SC values. None of them led to a successful phase-problem solution. These results illustrate the importance of screening for an optimal SC value, in some cases distinguishing success and failure in experimental phasing.

3.7. SIRAS: helicase-like transcription factor

The HIRAN domain of the helicase-like transcription factor was solved by SIRAS using Se as the heavy atoms (Kile *et al.*, 2015). The deposited structure (PDB ID 4s0n; Kile *et al.*,

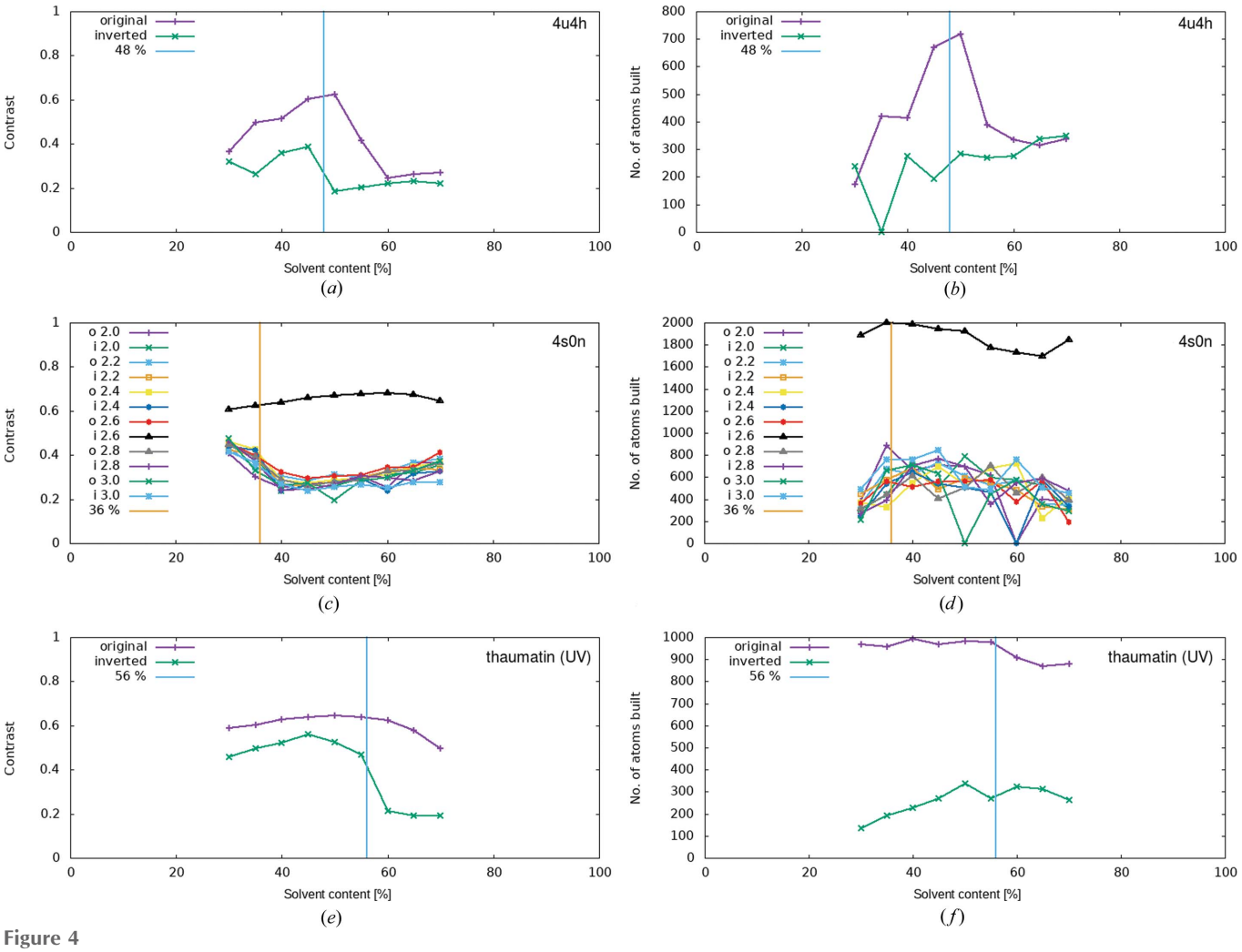


Figure 4
Characteristics of successful experimental phasing. N-terminal domain of UL21: (*a*), (*b*) the main characteristics of a successful main-chain tracing procedure using *SHELXE* in a narrow range of SC values. Helicase-like transcription factor: (*c*), (*d*) combination of *SHELXE* procedure results. Both solution hands (o = original, i = inverted) are shown for various phasing resolutions (numbers in Å). Thaumatin (macromolecular-crystallography tutorial, UV-RIP): (*e*), (*f*) results from the main-chain tracing procedure using *SHELXE* are similar to that of thaumatin structure determination using MAD. However, a minimal difference between contrasts belonging to the two hands is observed for an SC value that is optimal for the main-chain autotracing procedure at 40%. Vertical lines denote the real SCs of individual structures.

2015) also contains fragments of bound DNA. When all search parameters are known, the phase-problem solution is straightforward using the standard approaches as well as with *SHELIXIR*.

One of the very important parameters of phasing is the number of substructure atoms, which may frequently be unknown before the phase determination [*e.g.* soaking of selenourea (Luo, 2016) or halide ions (Dauter & Dauter, 2001)]. Simulation of such uncertainty resulted in an interesting observation in this case. At first, the resolution for phasing was automatically set to 1.9 Å and the number of heavy atoms was intentionally set to ten (true 16). Although the search for heavy-atom substructure was apparently successful for the limited number of heavy atoms in the space group $P2_1$, density modification and model tracing failed at such resolution. Screening for the optimal high-resolution limit was also performed with *SHELIXIR* using the same number of heavy atoms and resolution limits of 2.0, 2.2, 2.4, 2.6, 2.8 and 3.0 Å for the *SHELXD* routine. The heavy-atom substructures were nearly equivalent as analysed with the *phenix.emma* tool (r.m.s.d. of coordinates of the heavy atom sets lower than 0.3 Å). Interestingly, only the trials at 2.6 Å resolution led to successful phasing with *SHELXE*, while the use of the full set of heavy atoms led to success at other resolutions, including 1.9 Å. When the number of heavy atoms is unknown, we recommend searching for a significantly higher number [see heme PAS sensor domain of Ec DOS; Fig. 2(*b*)]. The results from the *SHELXE* routine are shown in Figs. 4(*c*) and 4(*d*).

To explore the influence of the high-resolution limit on the optimal SC value, we ran the same screening procedure with a search for 20 heavy atoms of the substructure. Phasing was successful at all resolutions. However, the optimal SC value varied from 35 to 50% with no observable trend related to the high-resolution limit for phasing (see Table S1 in the supporting information). This observation further underlines the importance of the screening approach implemented in *SHELIXIR*.

3.8. UV-RIP: thaumatin

RIP is used in a minority of experiments because of many technical difficulties, *e.g.* minimization of solvent around the crystal mounted on the loop, optimization of absorbed dose and high resolution in the case of UV-RIP. Nevertheless, RIP is a powerful tool because the data can be collected from the same crystal at the same position (de Sanctis *et al.*, 2016). Such an approach can also be used in synchrotron serial crystallography (Foos *et al.*, 2018).

Thaumatin was also used as a model protein for UV-RIP phasing in the aforementioned tutorial for learning and teaching macromolecular crystallography (Faust *et al.*, 2010). Similarly to the previous thaumatin data, a successful solution was found in space group $P4_1$ and in the correct space group $P4_12_12$. Here, a small difference is observed between the contrasts of the two hands at the optimal SC value of 40% [see Figs. 4(*e*) and 4(*f*)]. However, the remarkable difference in the number of atoms built with the *SHELXE* procedure indicates successful phasing.

The two thaumatin data sets (MAD, UV-RIP) share similar features in SC analysis. In both cases, the optimal SC value was 40%, which was also the maximal difference (16%) from the real SC of the above structures.

4. Technical limitations

Even if in our hands all the data sets selected for this study could be solved, we expect that in difficult cases alternative approaches or more sophisticated combinations of software will have to be used to determine the structure. *SHELIXIR* does not provide any guidance in the cases where 'successful' solutions are found in more than one space group. The user must evaluate such results individually and make his/her own decision on the correct solution. Users are encouraged to further prove the correct space group with refinement or use other automated tools such as *Zanuda* (Lebedev & Isupov, 2014). However, this does not affect the decision on whether the structure can be solved or not. Besides these limitations, automatic readout of the unit-cell parameters may fail. This is usually caused by nonstandard content of input files. In these cases, it is recommended to define the unit-cell parameters and the wavelength manually.

5. Conclusions

Experimental phasing does not usually represent a straightforward way to determine novel crystal structures. Successful phase-problem solution depends on a number of parameters. *SHELIXIR* is a tool that allows wide-scale computational screening that can be parallelized to some extent. All crystallographic routines are calculated with the *SHELXC*/*D*/*E* programs. Our tool offers simplicity together with minimal software dependencies. This approach has certain limitations, but it is highly suitable for rapid on-site data processing at beamlines and may be used to get a good starting point for more complicated data sets. As demonstrated above, the program could be used for phase-problem solution in routine data collection and analyses in projects focused on high-throughput structure determination, *e.g.* screening of novel targets for the treatment of diseases, structural genomics projects *etc.*

We tested the program on nine real cases that were reported previously. We selected the data to cover the whole phasing capability of the programs *SHELXC*/*D*/*E*. In all the reported cases, we imitated the situation of unknown space group, unknown number of heavy atoms, unknown protein content, unknown SC and unknown resolution range.

Modern data-reduction and evaluation tools, *e.g. XDS*, *phenix.xtriage* (Adams *et al.*, 2010) and *Aimless* (Evans & Murshudov, 2013), are able to suggest the correct space group with low ambiguity. However, for difficult cases, it may be beneficial to screen all space groups belonging to the determined Bravais type to narrow down the candidates, then build and refine the structure in the selected space groups

afterwards, and finally make the decision about the correct solution. Such an approach is enabled at the stage of phasing with *SHELIXIR*. Space groups to be tested, given by the Bravais types, are sorted according to their space group number. However, a specified order may be forced by the user on the basis of a previous data analysis, experience or expectations. Live monitoring of the calculation output on the HTML page may serve to further accelerate the search.

The SC value is expected to be one of the most critical parameters of the *SHELXE* routine for successful phase-problem solution. Screening for the optimum together with the poly-Ala main-chain autotracing procedure in *SHELXE* does not require *a priori* knowledge of the asymmetric unit composition. Our collection of examples shows that the optimal SC value does not necessarily have to match the real SC calculated after the crystal structure is solved and the composition is known. In some cases, better results (as judged by the number of atoms automatically built and refined in several cases) are obtained for SC values that differ by more than 10% from the real SC. Parallel execution of *SHELXE* procedures using *SHELIXIR* allows rapid screening of the SC value, which may provide improved initial phases and initial automated model building at a traditionally unexpected parameter value.

The benefit of optimization of the high-resolution limit for phasing has been reported previously and is already incorporated into several tools such as the *Auto-Rickshaw* (Panjikar *et al.*, 2009), *Crank2* (Winn *et al.*, 2011) and *Phenix* program packages. However, searching for an optimal low-resolution limit may be crucial in some cases as well. In one of our previous structures that was solved by molecular replacement (mouse NKR-P1a, PDB ID 3m9z; Kolenko *et al.*, 2011), the low-resolution limit had to be reduced down to 15 Å because of the presence of secondary weak diffraction coming from a smaller crystal that grew together with the main crystal in a different orientation. Overlap of the two lattices was shown to reduce the quality of the data. Moreover, in more favourable cases, optimizing the low-resolution cutoff can improve the output parameters (data not shown). Although we do not report the application of such an approach here, this function has been built into *SHELIXIR* and is available to users with similar or different data pathology in the low resolution that may or may not be apparent from visual analysis of raw images.

Acknowledgements

We thank Martin Dráb (Faculty of Nuclear Sciences and Physical Engineering, Czech Technical University in Prague) and Michal Strnad (Institute of Biotechnology of the Czech Academy of Sciences, BIOCEV) for providing IT support.

Funding information

This work was supported by the Ministry of Education, Youth and Sports CR and by the European Regional Development Fund (grant No. CZ.02.1.01/0.0/0.0/16_019/0000778 to the Faculty of Nuclear Sciences and Physical Engineering, Czech Technical University in Prague; grant No. CZ.02.1.01/0.0/0.0/15_003/0000447 to the Institute of Biotechnology of the Czech Academy of Sciences; grant No. CZ.1.05/1.1.00/02.0109 to the project BIOCEV; grant No. LM2018127 to the Institute of Biotechnology of the Czech Academy of Sciences, BIOCEV, CF Diffraction Techniques, part of Instruct-ERIC), the Czech Academy of Sciences (grant No. 86652036 to the Institute of Biotechnology of the Czech Academy of Sciences) and the Grant Agency of the Czech Technical University in Prague (grant No. SGS19/189/OHK4/3T/14 to the Faculty of Nuclear Sciences and Physical Engineering, Czech Technical University in Prague).

References

Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C. & Zwart, P. H. (2010). *Acta Cryst.* D**66**, 213–221.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). *Nucleic Acids Res.* **28**, 235–242.

Dauter, Z. & Dauter, M. (2001). *Structure*, **9**, R21–R26.

Ducros, V. M., Lewis, R. J., Verma, C. S., Dodson, E. J., Leonard, G., Turkenburg, J. P., Murshudov, G. N., Wilkinson, A. J. & Brannigan, J. A. (2001). *J. Mol. Biol.* **306**, 759–771

Evans, P. R. & Murshudov, G. N. (2013). *Acta Cryst.* D**69**, 1204–1214.

Faust, A., Puehringer, S., Darowski, N., Panjikar, S., Diederichs, K., Mueller, U. & Weiss, M. S. (2010). *J. Appl. Cryst.* **43**, 1230–1237.

Fisher, S. J., Levik, K. E., Williams, M. A., Ashton, A. W. & McAuley, K. E. (2015). *J. Appl. Cryst.* **48**, 927–932.

Foos, N., Seuring, C., Schubert, R., Burkhardt, A., Svensson, O., Meents, A., Chapman, H. N. & Nanao, M. H. (2018). *Acta Cryst.* D**74**, 366–378.

Grabowski, M., Langner, K. M., Cymborowski, M., Porebski, P. J., Sroka, P., Zheng, H., Cooper, D. R., Zimmerman, M. D., Elsliger, M.-A., Burley, S. K. & Minor, W. (2016). *Acta Cryst.* D**72**, 1181–1193.

Hendrickson, W. A., Smith, J. L. & Sheriff, S. (1985). *Methods Enzymol.* **115**, 41–55.

Kabsch, W. (2010). *Acta Cryst.* D**66**, 125–132.

Kile, A. C., Chavez, D. A., Bacal, J., Eldirany, S., Korzhnev, D. M., Bezsonova, I., Eichman, B. F. & Cimprich, K. A. (2015). *Mol. Cell*, **58**, 1090–1100.

Kolenko, P., Rozbeský, D., Vaněk, O., Kopecký, V., Hofbauerová, K., Novák, P., Pompach, P., Hašek, J., Skálová, T., Bezouška, K. & Dohnálek, J. (2011). *J. Struct. Biol.* **175**, 434–441.

Kurokawa, H., Lee, D. S., Watanabe, M., Sagami, I., Mikami, B., Raman, C. S. & Shimizu, T. (2004). *J. Biol. Chem.* **279**, 20186–20193.

Langer, G., Cohen, S. X., Lamzin, V. S. & Perrakis, A. (2008). *Nat. Protoc.* **3**, 1171–1179.

Lebedev, A. A. & Isupov, M. N. (2014). *Acta Cryst.* D**70**, 2430–2443.

Luo, Z. (2016). *Sci. Rep.* **6**, 37123.

McNicholas, S., Potterton, E., Wilson, K. S. & Noble, M. E. M. (2011). *Acta Cryst.* D**67**, 386–394.

Metrick, C. M., Chadha, P. & Heldwein, E. E. (2015). *J. Virol.* **89**, 2979–2984

Meyer, P. A., Socias, S., Key, J., Ransey, E., Tjon, E. C., Buschiazzo, A., Lei, M., Botka, C., Withrow, J., Neau, D., Rajashankar, K., Anderson, K. S., Baxter, R. H., Blacklow, S. C., Boggon, T. J., Bonvin, A. M. J. J., Borek, D., Brett, T. J., Caflisch, A., Chang, C. I., Chazin, W. J., Corbett, K. D., Cosgrove, M. S., Crosson, S., Dhe-Paganon, S., Di Cera, E., Drennan, C. L., Eck, M. J., Eichman, B. F., Fan, Q. R., Ferré-D'Amaré, A. R., Christopher Fromme, J., Garcia, K. C., Gaudet, R., Gong, P., Harrison, S. C., Heldwein, E. E., Jia, Z.,

Keenan, R. J., Kruse, A. C., Kvansakul, M., McLellan, J. S., Modis, Y., Nam, Y., Otwinowski, Z., Pai, E. F., Pereira, P. J. B., Petosa, C., Raman, C. S., Rapoport, T. A., Roll-Mecak, A., Rosen, M. K., Rudenko, G., Schlessinger, J., Schwartz, T. U., Shamoo, Y., Sondermann, H., Tao, Y. J., Tolia, N. H., Tsodikov, O. V., Westover, K. D., Wu, H., Foster, I., Fraser, J. S., Maia, F. R. N. C., Gonen, T., Kirchhausen, T., Diederichs, K., Crosas, M. & Sliz, P. (2016). *Nat. Commun.* **7**, 10882.

Panjikar, S., Parthasarathy, V., Lamzin, V. S., Weiss, M. S. & Tucker, P. A. (2009). *Acta Cryst.* D**65**, 1089–1097.

Pape, T. & Schneider, T. R. (2004). *J. Appl. Cryst.* **37**, 843–844.

Sanctis, D. de, Zubieta, C., Felisaz, F., Caserotto, H. & Nanao, M. H. (2016). *Acta Cryst.* D**72**, 395–402.

Sheldrick, G. M. (2008). *Acta Cryst.* A**64**, 112–122.

Sheldrick, G. M. (2010). *Acta Cryst.* D**66**, 479–485.

Sheldrick, G. M., Gilmore, C. J., Hauptman, H. A., Weeks, C. M., Miller, M. & Usón, I. (2012). *International Tables for Crystallography*, Vol. F, *Crystallography of Biological Macromolecules*, 2nd online ed., edited by E. Arnold, D. M. Himmel & M. G. Rossmann, pp. 413–429. Chester: International Union of Crystallography.

Sonnenberg, A., Rojas, A. M. & de Pereda, J. M. (2007). *J. Mol. Biol.* **368**, 1379–1391.

Usón, I. & Sheldrick, G. M. (2018). *Acta Cryst.* D**74**, 106–116.

Vollmar, M., Parkhurst, J. M., Jaques, D., Baslé, A., Murshudov, G. N., Waterman, D. G. & Evans, G. (2020). *IUCrJ*, **7**, 342–354.

Vonrhein, C., Blanc, E., Roversi, P. & Bricogne, G. (2007). *Methods Mol. Biol.* **364**, 215–230.

Vonrhein, C., Flensburg, C., Keller, P., Sharff, A., Smart, O., Paciorek, W., Womack, T. & Bricogne, G. (2011). *Acta Cryst.* D**67**, 293–302.

Weeks, C. M., Adams, P. D., Berendzen, J., Brunger, A. T., Dodson, E. J., Grosse-Kunstleve, R. W., Schneider, T. R., Sheldrick, G. M., Terwilliger, T. C., Turkenburg, M. G. W. & Usón, I. (2003). *Methods Enzymol.* **374**, 37–83.

Weinert, T., Olieric, V., Waltersperger, S., Panepucci, E., Chen, L., Zhang, H., Zhou, D., Rose, J., Ebihara, A., Kuramitsu, S., Li, D., Howe, N., Schnapp, G., Pautsch, A., Bargsten, K., Prota, A. E., Surana, P., Kottur, J., Nair, D. T., Basilico, F., Cecatiello, V., Pasqualato, S., Boland, A., Weichenrieder, O., Wang, B. C., Steinmetz, M. O., Caffrey, M. & Wang, M. (2015). *Nat. Methods*, **12**, 131–133.

Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G. W., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A. & Wilson, K. S. (2011). *Acta Cryst.* D**67**, 235–242.

Part 3)

Studies of biotechnological proteins: Structure-function analysis of tetracycline-modifying enzyme from *Stenotrophomonas maltophilia*

3.1 Materials and methods

3.1.1 Bioinformatic analysis – target selection

A search for the nucleotide sequences coding for protein in *Stenotrophomonas maltophilia* was performed using the NCBI database (NCBI Resource Coordinators, 2016) with the *Protein BLAST* (Boratyn *et al.*, 2013) with a help of analysis tools from EMBL-EBI (Madeira *et al.*, 2019, 2022). The sequence of the putative tetracycline-modifying enzyme from *S. maltophilia* strain AB550 (*Sm*TetX) with the NCBI reference code WP_049406473 (Arita *et al.*, 2021; O'Leary *et al.*, 2016) from the GenBank entry CP028899 (Glady-Croue *et al.*, 2018) was selected as a target protein for investigation (Figure 8). The presence of the nucleotide sequence coding for this protein was reported in clinical isolates from Australia, India, the USA (Pak *et al.*, 2015; Roach *et al.*, 2015) and Europe (Esposito *et al.*, 2017; Lira *et al.*, 2017). Initial structure modelling trials were carried out with the servers *I-TASSER* (Roy *et al.*, 2010; Yang *et al.*, 2015; Zhang, 2008) and *SWISS-MODEL* (Waterhouse *et al.*, 2018).

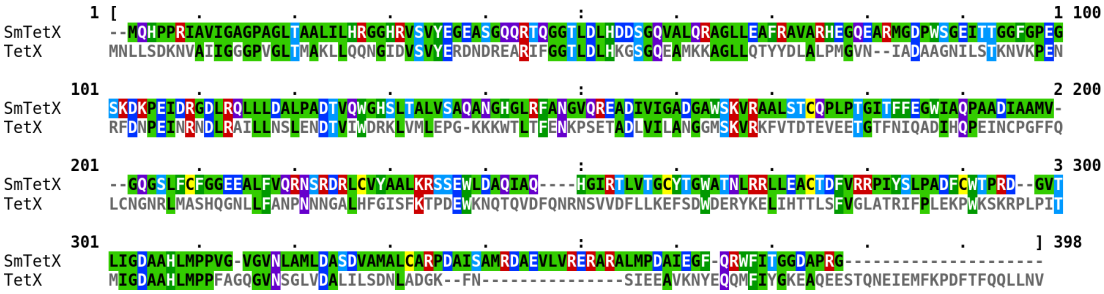


Figure 8: Sequence alignment of *Sm*TetX and TetX from *Bacteroides thetaiotaomicron* (Volkers *et al.*, 2011; PDB entry 2Y6R), carried out with *Clustal Omega* (Sievers *et al.*, 2011) at EMBL-EBI (Goujon *et al.*, 2010; McWilliam *et al.*, 2013). The sequence identity is 27.6 %. The graphics was created with *MView* (Madeira *et al.*, 2022).

3.1.2 Recombinant expression

After back-translation of the target amino-acid sequence WP_049406473, tools *OPTIMIZER* (Puigbò *et al.*, 2007) and *NEBcutter* (Vincze *et al.*, 2003) were used to optimize the codon usage for *Escherichia coli* and to inspect restriction sites in the sequence. The cleavage site for Tobacco etch virus (TEV) protease was placed between the 6xHis-Tag and the target *Sm*TetX sequence. Thus, the amino-acid sequence of the protein construct starts with MGSSHHHHHHSSGENLYFQGH and is followed by the first amino acid of sequence WP_049406473. The plasmid encoding the described construct was synthesized and cloned into the pET28a+TEV expression vector by GenScript (USA); restriction sites *NdeI* and *BamHI* were used.

With the heat shock method, the vector was transformed to competent *Eschericia coli* cells, strain Lemo21 (DE3). Cell precultures grew in LB Broth medium in a shaker at 32 °C and 180 rpm overnight. Antibiotics (50 µg/ml kanamycin and 25 µg/ml chloramphenicol) were added to media. Then, one l of Power Broth™ medium (Molecular Dimensions catalog No MD12-106) was supplemented with 10 ml of the preculture and this cell culture was subsequently incubated in a shaker at 30 °C and 150 rpm for three hours. When OD600 ~ 0.5, the culture was supplemented with 1 mM isopropyl β-d-1-thiogalactopyranoside to induce expression. After 16 hours in a shaker at 20 °C and 150 rpm, the culture was centrifuged at 5000g for 30 min.

3.1.3 Purification

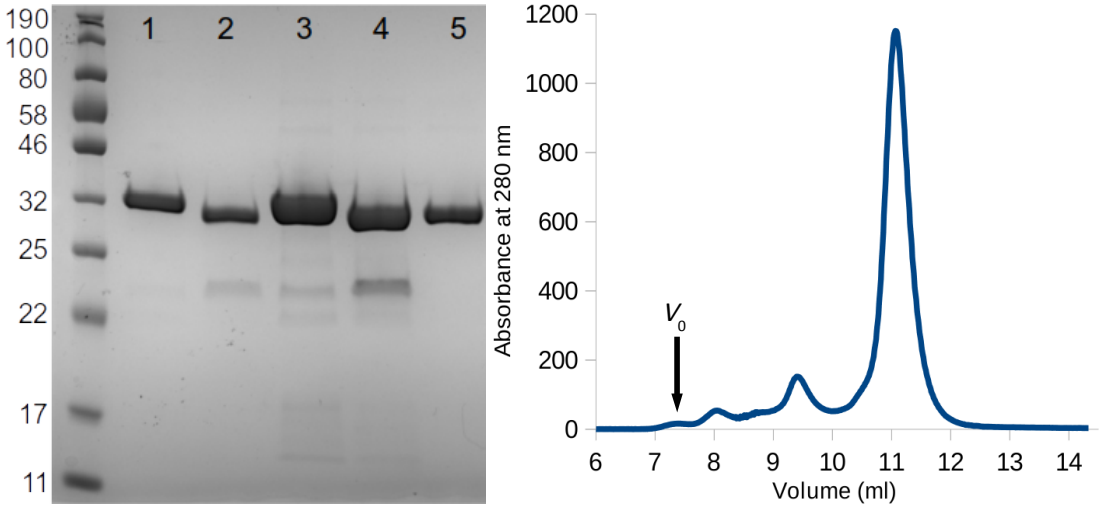
Composition of buffers used in this section:

- Cell lysis buffer: 50 mM Tris-HCl, 500 mM NaCl, 30 mM imidazole, pH 8, DNase I (Sigma-Aldrich catalog No D4263), protease inhibitor cocktail (Sigma-Aldrich catalog No P8849).
- Wash buffer: 50 mM Tris-HCl, 500 mM NaCl, 30 mM imidazole, pH 8.
- Elution buffer: 50 mM Tris-HCl, 500 mM NaCl and 500 mM imidazole, pH 8.
- TEV protease reaction buffer: 50 mM Tris-HCl, 500 mM NaCl, 0.5 mM dithiothreitol (DTT), 1 mM ethylenediaminetetraacetic acid (EDTA), pH 8.
- Size-exclusion chromatography (SEC) buffer: 25 mM Tris-HCl, 150 mM NaCl, pH 8.

The cell paste was mixed and dissolved in the 'lysis buffer'. Cell lysis was carried out gently using a sonicator Qsonica Q700. The process with an amplitude of 1 took 10 minutes of operation in total, two seconds in operation was followed by four seconds of pause. The lysate was centrifuged for 30 minutes at 40000g and filtred.

Subsequently, affinity chromatography was performed using an ÄKTA purifier (GE Healthcare/Amersham Biosciences) at 10 °C. The sample was loaded onto a charged 5 ml Ni-NTA HisTrap™ FF column (GE Healthcare) and washed with 50 ml of the 'wash buffer'. The protein was eluted with a stepped gradient of imidazole (30 % of the 'elution buffer', *i.e.* the concentration of imidazole was 171 mM). After a transfer into the 'TEV protease reaction buffer' using a 3 kDa cut-off Nanosep centrifugal device (Pall Corporation), the TEV cleavage was carried out at 4 °C overnight using the TEV-protease produced on the basis of the regular protocol (Tropea *et al.*, 2009). The subsequent separation of the tag-free protein was done with the 5 ml Ni-NTA HisTrap™ FF column (GE Healthcare). The result from a sodium dodecyl sulfate-polyacrylamide gel electrophoresis is shown in Figure 9a.

To obtain a sample of very high purity, the protein was finally purified using size-exclusion chromatography (SEC) conducted on an ÄKTA purifier (GE Healthcare/Amersham Biosciences). The column Superdex 75 Increase 10/300 GL (GE Healthcare) was equilibrated with the 'SEC buffer'. Protein dilution was monitored with absorbance at 280 nm (Figure 9b).



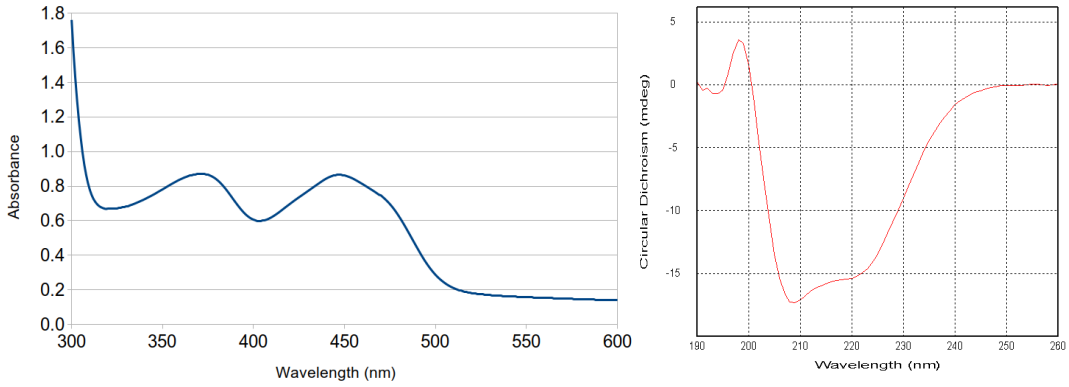
(a) (b)

Figure 9: (a) Sodium dodecyl sulfate-polyacrylamide gel electrophoresis of *Sm*TetX showed that the cleavage of 6xHis-tag with the TEV-protease was successful. Sample 1: *Sm*TetX (3 µg) before the tag cleavage; sample 2: *Sm*TetX (3 µg) after the tag cleavage; sample 3: *Sm*TetX (10 µg) before the tag cleavage; sample 4: *Sm*TetX (10 µg) after the tag cleavage; sample 5: *Sm*TetX (3 µg) after the tag cleavage and separation using Ni-NTA. The gel using a marker Color Prestained Protein Standard, Broad Range marker (New England Biolabs) in the left lane was stained with InstantBlue (Expedeon). The gel was prepared from Bio-Rad TGX™ FastCast™ Acrylamide Kit 12%. (b) Chromatogram of size-exclusion chromatography of the *Sm*TetX sample after the His-tag cleavage with TEV-protease. The pure target protein was eluted after a running volume of 10 ml. The void volume (V_0 = 7.3 ml) is marked with an arrow.

3.1.4 Protein sample characterization

The *Sm*TetX spectrum in the UV-VIS range (Figure 10a) was measured in a Specord 50 Plus spectrometer (Analytika Jena). The protein at a concentration of 3 mg ml^{-1} was stored in 100 mM TAPS, pH 8.5.

The circular dichroism spectrum was measured using a Chirascan Plus spectrometer (Applied Photophysics). A protein sample of *Sm*TetX with a concentration of 0.3 mg ml^{-1} was placed in a 1 mm cell. The result showed that *Sm*TetX was folded properly (Figure 10b).



(a) (b)

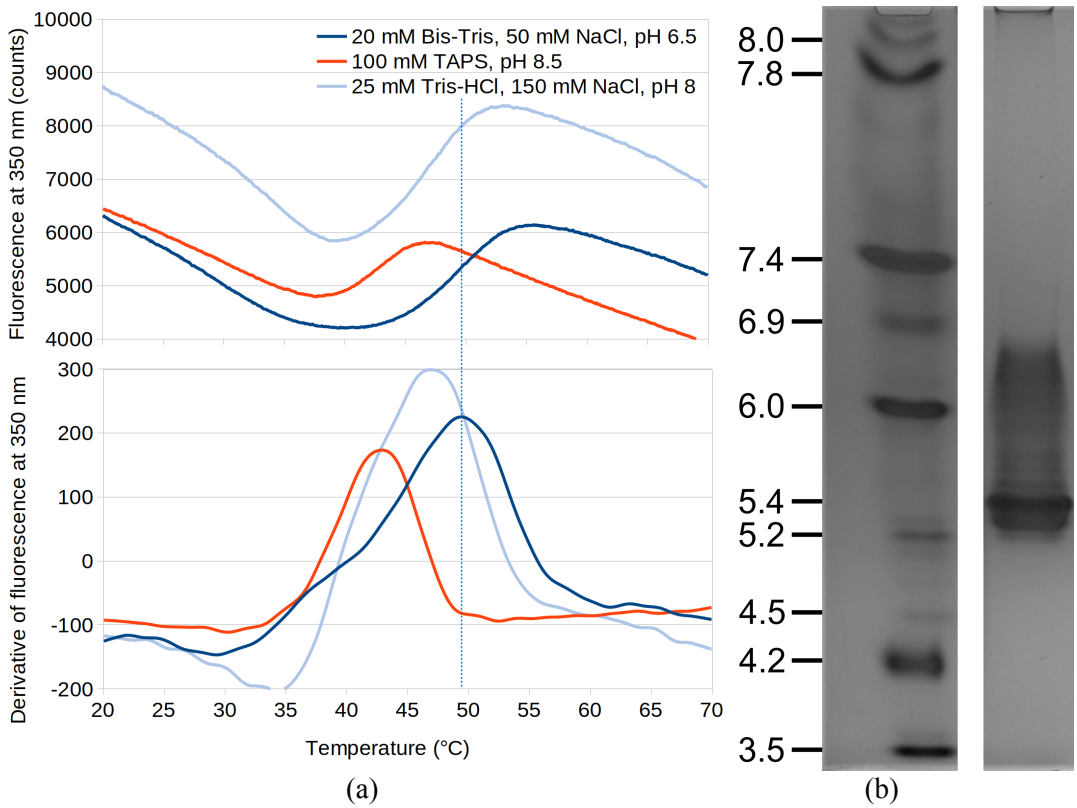
Figure 10: UV-VIS spectrum of *Sm*TetX shows two characteristic peaks for the oxidized state of FAD, however, a partial presence of the reduced state is also possible owing to the equal heights of both peaks.

Protein stability was validated with nanoscale differential scanning fluorimetry. Screening of several buffers in a broad pH-range was performed using a Prometheus NT.48 (NanoTemper). The highest melting temperature of the protein (Figure 11a) was achieved in the buffer consisting of 20 mM Bis-Tris, 50 mM NaCl, pH 6.5: 49.4 °C. Thus, this buffer was then used for storage and following experiments.

Isoelectric focusing was used to determine the isoelectric point experimentally. On a 5 % polyacrylamide gel with a fixed vertical pH gradient (Novex pH 3-10 IEF, ThermoFisher Scientific), *Sm*TetX (10 µg) was loaded. The corresponding marker IEF Marker 3-10 (ThermoFisher Scientific) was used. The electrophoresis ran with a voltage of 100 V for 60 min, then 200 V for 60 min, and finally 500 V for 30 min. The gel fixation was done with 12 % trichloracetic acid for 30 min and the staining with InstantBlue (Expedeon). The isoelectric point of *Sm*TetX is 5.4 (Figure 11b) which is in good agreement with the prediction (5.7) from *PROTPARAM* (Gasteiger *et al.*, 2005).

3.1.5 Activity assay

The spectrophotometric assay was carried out in a similar setup as reported previously (Forsberg *et al.*, 2015; Gasparrini *et al.*, 2020; Moore *et al.*, 2005; Rudra *et al.*, 2018; Yang *et al.*, 2004). The measurement was done in 96-well plates (BRAND, Wertheim, GE) in triplicates at 25 °C in a total volume of 100 µl. Individual enzymatic reactions contained 0.1 µM *Sm*TetX enzyme in 100 mM TAPS buffer (pH 8.5) and the following reagents: 0.5 mM oxytetracycline (OTC) and 0.5 mM NADPH. The experiment was monitored in a selected UV-VIS range every 15 min in a Spark (TECAN) microplate reader. The OTC consumption was monitored at 400 nm due to the overlap of OTC and NADPH spectra that possess maxima at 368 nm and 340 nm, respectively.



(a) (b)

Figure 11: (a) Nanoscale differential scanning fluorimetry monitored at 350 nm using selected buffers. The buffer composed of Bis-Tris and NaCl (pH 6.5) stabilized the protein. The corresponding melting temperature of 49.4 °C is marked as a dashed line. (b) Isoelectric focusing gel electrophoresis determined the isoelectric point of the protein: 5.4. The gel lanes between the marker (ThermoFisher Scientific IEF Marker 3-10) on the left and the *Sm*TetX sample were cut off the graphics without detriment to the result accuracy.

3.1.6 Mass spectrometry

Contribution note: the experiments described in this section were carried out by Petr Pompach from Centre of Molecular Structure, Institute of Biotechnology of the Czech Academy of Sciences.

After alkylation of *Sm*TetX with 30 mM iodoacetamide and incubation for 30 min in dark, trypsin (0.1 µg/ml) was added and the enzymatic reaction was incubated for 16 hours at 37 °C.

The experiment was conducted using the following setup: an Agilent 1200 series liquid chromatography system (Agilent Technologies) connected to a timsToF Pro PASEF mass spectrometer, set to standard proteomics PASEF method, combined with a Captive spray (Bruker Daltonics) in positive data-dependent mode operation.

The peptide mixture (5 µl) was injected on a UHPLC Fully Porous Polar C18 2.1mm ID column (Phenomenex) heated at 50 ˚C, followed by trapping with a flow rate of 20 µl/min for 5 min. Subsequently, separation of peptides was carried out with a flow rate of 4 µl/min on a Luna Omega 3 µm Polar C18 100 Å 150x0.3 mm column

(Phenomenex) heated at 50 ˚C using a linear gradient of acetonitrile, 5 % to 35 % (v/v).

The measured data were processed with PeaksStudio 10.0 software (Bioinformatics Solutions, Canada). The data were searched against the *Sm*TetX sequence using the following parameters: enzyme – trypsin (specific), carbamidomethylation and oxidation of methionine as variable modifications.

Except for the first six residues, the sequence of *Sm*TetX is fully covered. The results showed that all the cysteines of *Sm*TetX were alkylated (Figure 12). Thus, the protein in solution does not possess any disulfide bridges.

Figure 12: The *Sm*TetX sequence is fully covered with peptides detected using mass spectrometry. The observed alkylation for all the cysteines in sequence implies no cysteine bridges are involved in the protein structure.

3.1.7 Small-angle X-ray scattering

The buffer of the most concentrated fraction of protein from the size-exclusion chromatography was changed to the storage buffer (25 mM Bis-Tris, 150 mM NaCl, pH 6.5) via dialysis in Slide-A-Lyzer MINI Dialysis Devices, 3.5K MWCO (Thermo Scientific). The obtained sample was centrifuged at 15 000g for 15 min and measured with the small-angle X-ray scattering (SAXS) method at the Centre of Molecular Structure (Institute of Biotechnology AS CR, Czech Republic). The protein concentration in the samples was determined with a DS-11+ spectrophotometer (DeNovix).

Analysis was performed using two samples: a control sample without any modification and a sample with reducing agent dithiothreitol (DTT), added to a concentration of 30 mM. The data were collected using an Anton Paar SAXSpoint 2.0 equipped with an Excillum MetalJet C2+ X-ray source with gallium anode (Hemberg, 2004; Hållstedt *et al.*, 2021) and a Dectris EIGER R 1M detector (Casanas *et al.*, 2016; Dinapoli *et al.*, 2011; Johnson *et al.*, 2014), the measurements were conducted in a vacuum. The two-dimensional scattering images were reduced to one-dimensional scattering profiles in *Aares* (J. Stránský, *unpublished*), optimized software for the in-house source. The buffer subtraction was performed in *PRIMUS* (Konarev *et al.*, 2003). Parameters of the experiments and results from further analysis with several programs from the *ATSAS* software package (Konarev *et al.*, 2003; Manalastas-Cantos *et al.*, 2021; Petoukhov *et al.*, 2012) are shown in Table 2.

The measured data and predicted curve based on the atomic structure of the *Sm*TetX monomer are plotted in Figure 13.

A partial aggregation of the sample is indicated with the data from very small angles. The Guinier analysis of both samples reveals an equal radius of gyration R_g = 2.36 nm (Figure 13c). Analysis using Kratky plots confirms a globular shape of the protein (Figure 13d). Both experimental data are well fitted with the monomeric model based on the crystal structure (described below), the agreements between the model and data are conclusive (χ^2 equals 0.90 and 1.05 for the sample with reducing agent and the control sample, respectively). In contrast, the dimeric model fits the data very poorly which is indicated also with the obtained values of χ^2 higher than 5. Hence, the protein is in solution in a monomeric state.

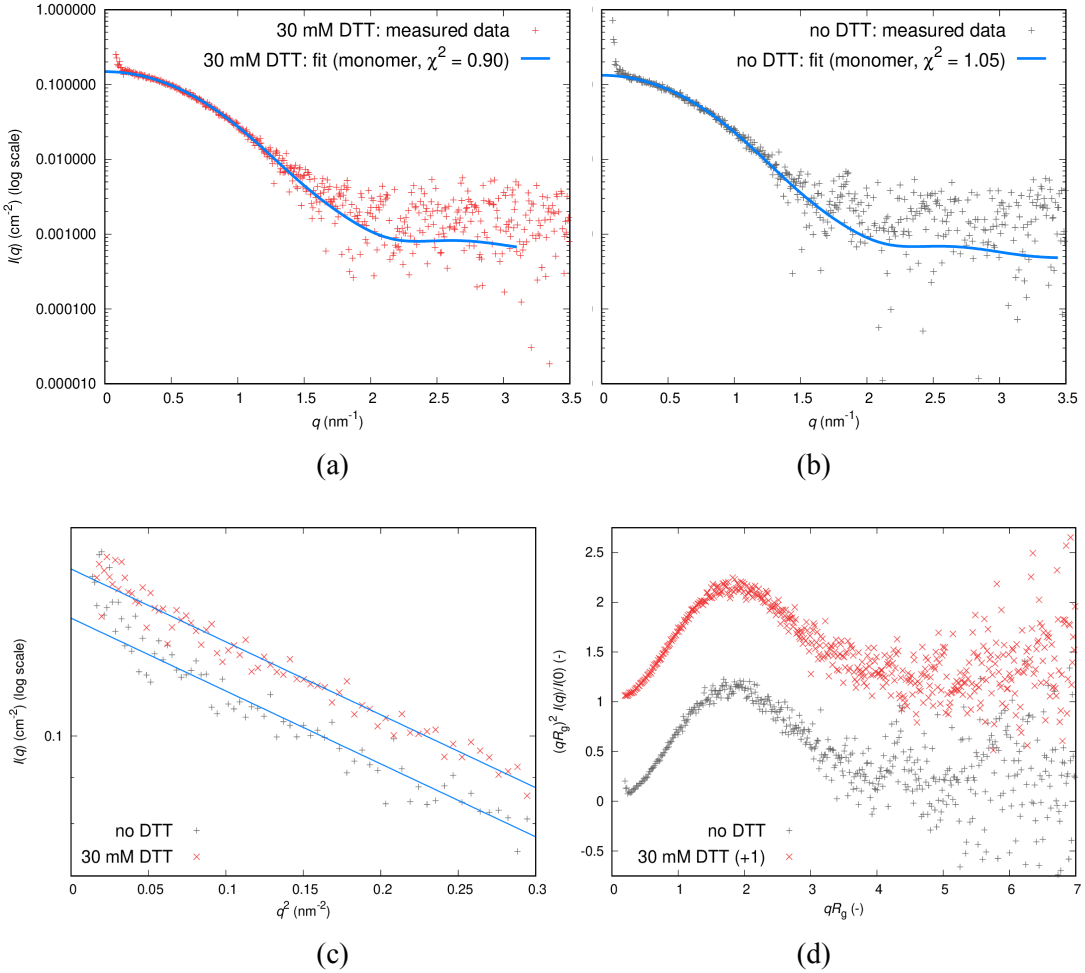


Figure 13: Small-angle X-ray scattering measurement of *Sm*TetX. (a-b) Intensities for the sample with (a, red) and without (b, gray) the reducing agent. The calculated curve of a monomer fits well both datasets (blue line). (c) The corresponding Guinier analysis with linear fits, colored as in (a-b). (d) Dimensionless Kratky plots, colored as in (a-b), exhibit a characteristic shape for globular proteins.

Table 2: Small-angle X-ray scattering parameters of data collection and processing for *Sm*TetX.

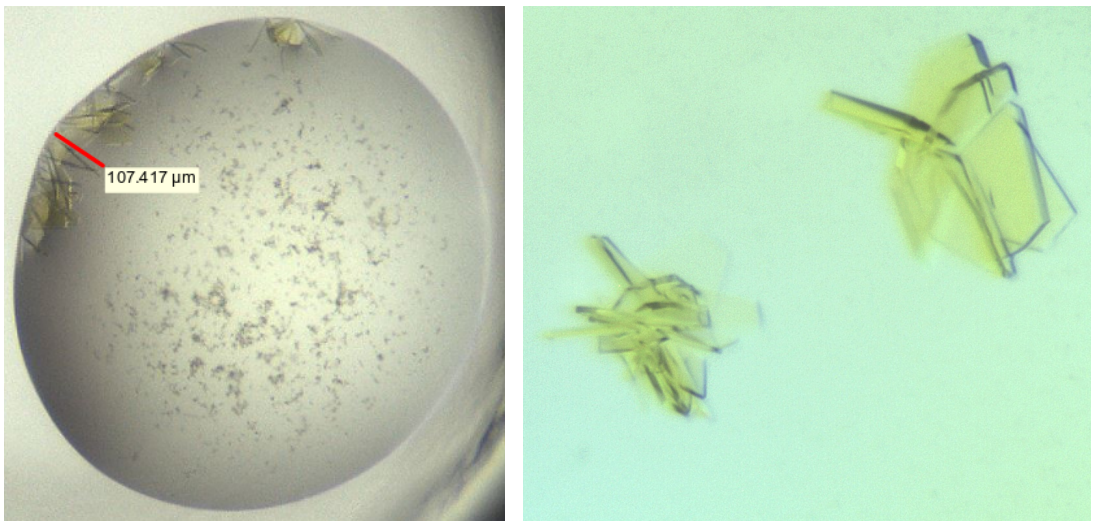
Sample details	with 30 mM DTT	without DTT
SASBDB accession code	SASDPW7	SASDPV7
Buffer	25 mM Bis-Tris, 150 mM NaCl, pH 6.5, 30 mM DTT	25 mM Bis-Tris 150 mM NaCl, pH 6.5
Theoretical M_W (Da)	38941	
Concentration (mg ml^{-1})	2.4	
SAXS data-collection parameters		
Instrument	Excillum MetalJet C2+ X-ray source with Anton Paar SAXSpoint 2.0 and Dectris EIGER R 1M detector	
X-ray wavelength (Å)	1.34	
Beam size at sample (mm)	0.94	
Beamstop size (mm)	2	
Sample to detector distance (mm)	826	
Measurement q range (nm^{-1})	0.07 - 4.14	
Sample temperature (°C)	4	
Exposure time (s)	90 x 15.6	61 x 15.6
Parameters of conversion to absolutes scale: Calibrated using water as scattering standard		
Guinier analysis (*AUTORG*) (Petoukhov *et al.*, 2007)		
R_g (nm)	2.36 ± 0.03	2.36 ± 0.04
$I(0)$ (cm^{-2})	0.153 ± 0.001	0.135 ± 0.001
Point index, of the Guinier region	9 - 82	18 - 82
q range (nm^{-1})	0.14 - 0.55	0.18 - 0.55
Fidelity	0.97	0.75
Aggregation index	0.03	0.25
Analysis using indirect Fourier transform (Svergun, 1992)		
R_g (nm)	2.28 ± 0.01	2.32 ± 0.02
$I(0)$ (cm^{-2})	0.1501 ± 0.0009	0.133 ± 0.001
q range (nm^{-1})	0.129 - 1.802	0.181 - 1.802
Real space range (nm)	0.0 - 6.4	0.0 - 6.8
Fit quality	0.79	0.74
Ambimeter	1.15	0.95
Shannon analysis (*SHANUM*) (Konarev and Svergun, 2015)		
Optimal Shannon channels	7	8
Optimal q_{max} (nm^{-1})	3.10	3.44
Molecular weight estimates:		
M_W from Porod Volume (Petoukhov *et al.*, 2012)		
V_P (nm^3)	22.57	18.19
M_W (Da)	14107	11371
M_W from Porod Invariant (Hajizadeh *et al.*, 2018)		
M_W (Da)	27961	35544
M_W from Volume of Correlation (Rambo and Tainer, 2013)		
V_C (nm^3)	309.86	319.07
M_W (Da)	33057	35057

Sample details	with 30 mM DTT	without DTT
M_W from Apparent Volume (Fischer et al. 2010)		
Q (nm^{-3})	0.275	0.257
V_{app} (nm^3)	71.77	76.76
M_W (Da)	35731	38559
M_w from Size and Shape (Franke *et al.*, 2017)		
M_W (Da)	37793	40146
M_W from Bayesian inference (Hajizadeh *et al.*, 2018)		
M_W (Da)	33100	36900
M_W score	0.28	0.47
Analysis using indirect Fourier transform (*GNOM*) (Svergun, 1992)		
R_g (nm)	2.28 ± 0.01	2.32 ± 0.02
$I(0)$ (cm^{-2})	0.1501 ± 0.0009	0.133 ± 0.001
q range (nm^{-1})	0.129 - 1.802	0.181 - 1.802
Real space range (nm)	0.0 - 6.4	0.0 - 6.8
Fit quality	0.79	0.74
Ambimeter	1.15	0.95
Agreement between data and atomic model (*CRYSOL*) (Manalastas-Cantos *et al.*, 2021)		
M_W from structure on input (Da)	37980	37980
q range (nm^{-1})	0.1 - 1.5	0.1 - 1.5
Predicted R_g (nm)	2.25	2.27
χ^2	0.90	1.05
Results of ab-initio shape determination		
Particle symmetry	*P*1	*P*1
Particle anisometry	unknown	unknown
q range (nm^{-1})	0.129 - 1.802	0.181 - 1.802
DAMMIF (Franke and Svergun, 2009)		
M_W (Da)	35000 ± 100	36700 ± 100
Particle R_g (nm)	2.2805 ± 0.0002	2.3201 ± 0.0002
Particle D_{max} (nm)	8.0 ± 0.1	9.1 ± 0.5
χ^2	0.8129 ± 0.0007	0.8541 ± 0.0003
Number of models on input	32	32
No. of excluded models	2	2
DAMMIN (Svergun, 1999)		
Particle R_g (nm)	2.29	2.33
Particle D_{max} (nm)	7.33	7.76
χ^2	0.80	0.84

3.1.8 Crystallization

The protein sample was transferred to the buffer consisting of 20 mM Bis-Tris, 50 mM NaCl, pH 6.5, using a 3 kDa cut-off Nanosep centrifugal device (Pall Corporation). This buffer showed the highest protein stability according to nanoscale differential scanning fluorimetry. The final protein concentration was 10 mg ml^{-1}. To investigate an optimal crystallization condition, several crystallization screens were set onto 96-well crystallization plates with a Gryphon crystallization robot (Art Robbins) using the sitting drop vapour diffusion method. The plates were monitored in a crystallization hotel RI1000 (Formulatrix) at 20 °C.

Small protein crystals (Figure 14a) grew in the MORPHEUS screen (Molecular Dimensions), the condition C2 [10% (w/v) polyethylene glycol 8000, 20% (v/v) ethylene glycol, 30 mM sodium nitrate, 30 mM disodium hydrogen phosphate, 30 mM ammonium sulfate, 100 mM MES/imidazole, pH 6.5] (Gorrec, 2009). The crystallization condition was further optimized in the hanging drop arrangement, including adjustment of precipitant concentration, microseeding and additive screening (Hampton Research). Crystals in the form of plates with a length of over 100 µm, further used for the X-ray analysis and shown in Figure 14b, were obtained in the optimized condition consisting of 12% (w/v) polyethylene glycol 8000, 24% (v/v) ethylene glycol, 60 mM sodium nitrate, 60 mM disodium hydrogen phosphate, 60 mM ammonium sulfate, 100 mM MES/imidazole pH 6.5 and 4% acetone. The drop was composed of 1 µl of the protein sample and 0.5 µl of the reservoir solution.



(a) (b)

Figure 14: Crystals of the *Sm*TetX. (a) First crystallization hit in the MORPHEUS screen, condition C2 (Gorrec, 2009). (b) Clusters of plate-shaped crystals grown in the optimized condition. These individual single crystals, after the manual division of the clusters, diffracted sufficiently in the X-ray beam.

3.1.9 Diffraction data collection and processing

Crystals were mounted from drops into LithoLoop loops (Molecular Dimensions). Vitrification in liquid nitrogen was carried out without prior soaking in a cryoprotectant solution as the crystallization condition contained ethylene glycol in a sufficient concentration. The diffraction experiment was performed at BESSY II synchrotron radiation source, beamline 14.1 (Helmholtz Zentrum Berlin, Germany). Data were collected using a Dectris PILATUS 6M detector (Kraft *et al.*, 2009; Loeliger *et al.*, 2012) and a mini kappa goniometer (Mueller *et al.*, 2015). The experiment was controlled and monitored with *MXCuBE* (Gabadinho *et al.*, 2010; Oscarsson *et al.*, 2019), *iMOSFLM* (Battye *et al.*, 2011; Leslie, 2006) and *XDSAPP* (Krug *et al.*, 2012; Sparta *et al.*, 2016).

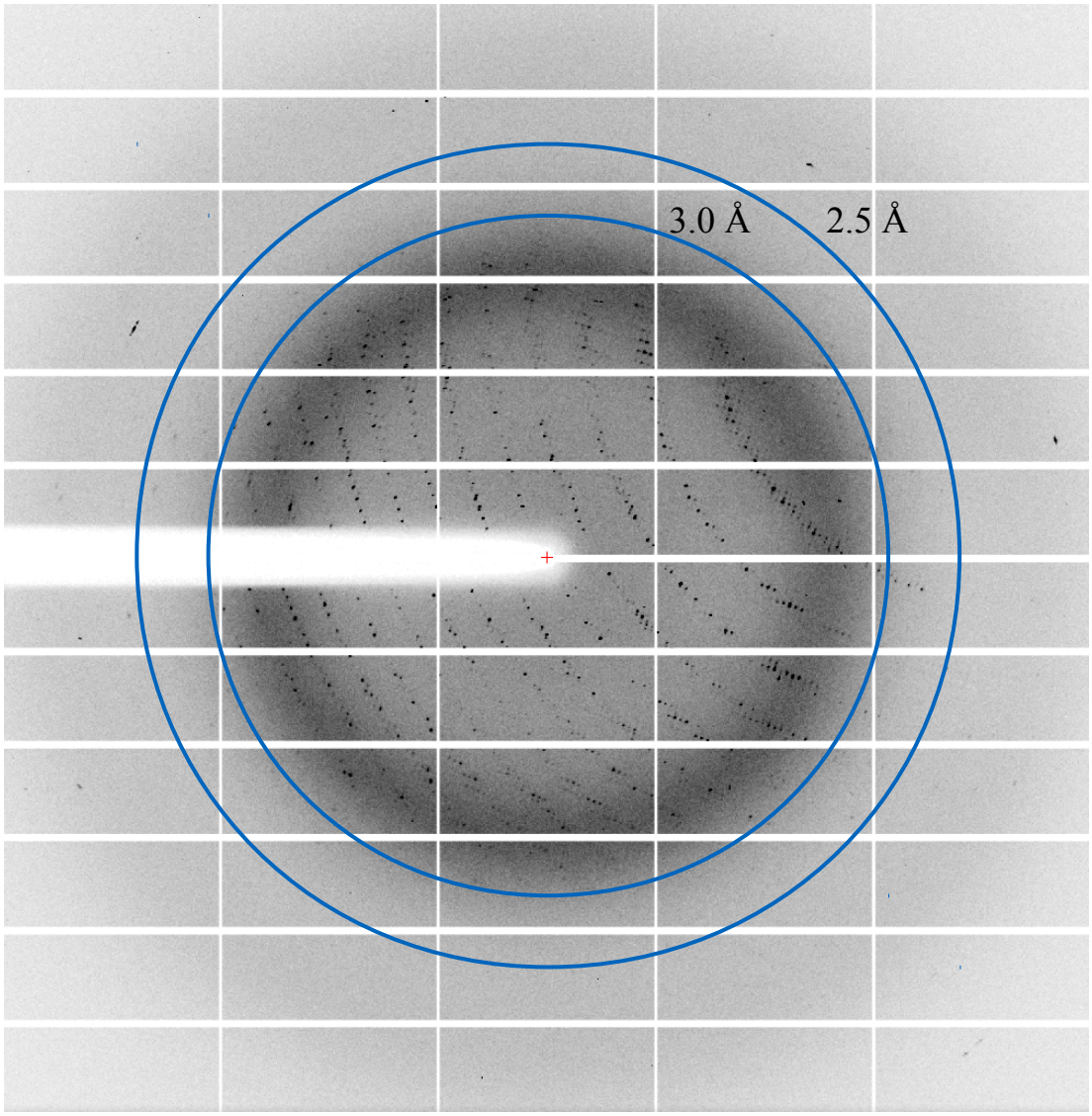


Figure 15: Diffraction image from the *Sm*TetX crystal characterization with an exposure time of 1 s and an oscillation of 1° shows severe anisotropy. Resolution levels of 2.5 Å and 3.0 Å are shown in blue. The graphics was created with *Adxv* (2022).

The reflections were indexed and integrated with *XDS* and *XDSgui* (Kabsch, 2010a, 2010b) and scaled in *AIMLESS* (Evans, 2011; Evans and Murshudov, 2013) from *CCP*4 (Winn *et al.*, 2011). Owing to the observed radiation damage to the crystal, diffraction images from the final stage of the data collection were not included in the processing. The measured data possess severe anisotropy that is visible on the diffraction images (Figure 15). While taking into account a criterion of $CC_{1/2}$ to be higher than 0.30, a high-resolution cutoff suggested by *AIMLESS* differed from 2.43 Å to 1.92 Å depending on the direction in the reciprocal unit cell. Thus, an anisotropy correction using the unmerged data from *XDS* was applied in *STARANISO* (Tickle *et al.*, 2018). The diffraction data statistics are shown in Table 3.

Table 3: Data collection parameters and merging statistics for *Sm*TetX after anisotropy correction. Values for the highest resolution shell are given in parentheses.

PDB entry	8AQ8
X-ray source	BESSY II, BL14.1
Wavelength (Å)	0.9180
Detector	PILATUS 6M
Temperature (K)	100
Crystal-to-detector distance (mm)	424
No. of oscillation images	2400
Exposure time per image (s)	0.1
Oscillation width (°)	0.1
Space group	$P2_1$
Cell parameters (Å, °)	$a = 52.9$ $b = 160.5$ $c = 95.6$ $\beta = 95.9$
Resolution (Å)	48.21-1.95 (2.01-1.95)
No. reflections	467,117 (25,410)
No. unique reflections	98,291 (4,915)
Multiplicity	4.8 (5.2)
Completeness (spherical, %)	85.3 (48.5)
Completeness (ellipsoidal, %)	96.3 (92.9)
R_{meas}	0.156 (1.702)
R_{pim}	0.070 (0.729)
Mean $I/\sigma(I)$	8.6 (1.1)
$CC_{1/2}$	0.997 (0.378)
CC* (Karplus and Diederichs, 2012)	0.999 (0.741)
Mosaicity (°)	0.2
Wilson *B*-factor ($Å^2$)	39

The phase problem was solved with the molecular replacement method carried out in *Phaser* (Evans and McCoy, 2008; McCoy *et al.*, 2007) using data at 2.4 Å resolution. As the template structure, the putative FAD-dependent reductase in the abyssomicin biosynthesis pathway (Clinger *et al.*, 2021; PDB entry 6N04) was used. According to the phasing pipeline *MoRDa* (Krissinel *et al.*, 2018; Vagin and Lebedev, 2015), the template structure was divided into two domains and manually modified, which led

to an initial model with the following statistics: *R*-value = 35.4; LLG = 16956. The asymmetric unit consists of four protein molecules possessing translational non-crystallographic symmetry (NCS). Assignment to the space group $P2_1$ was checked with *ZANUDA* (Lebedev and Isupov, 2014). The initial model was partially rebuilt with *PHENIX AutoBuild* (Terwilliger, *et al.*, 2008).

Refinement of the structure model was performed in *REFMAC*5 (Kovalevskiy *et al.*, 2018; Murshudov *et al.*, 2011) with the stereochemical restraints for the molecule of FAD, generated in *AceDRG* (Long *et al.*, 2017) and manually customized as described previously (Švecová *et al.*, 2021). *Coot* (Emsley *et al.*, 2010) was used for the real-space refinement and other manual modifications of the structure model.

To confirm the suitability of the anisotropic high-resolution cutoff, paired refinement was run with *PAIREF* (Karplus and Diederichs, 2012; Malý *et al.*, 2020, 2021) from 2.4 Å stepwise up to 1.95 Å. The overall R_{free}-values decreased with every added high-resolution shell that indicated improvement of the structure model (Table 4).

Table 4: Overall *R*-values from the paired refinement of the structure model of *Sm*TetX, run with *PAIREF* (Malý *et al.*, 2020). For each incremental resolution step *X* (init.) → *Y*(final), Δ*R* denotes the difference between the *R*-values of the models refined at the higher resolution *Y* and lower resolution *X*; the *R*-values were calculated at the lower resolution *X*. Results show continuous improvement of the model while the involvement of the high-resolution data up to 1.95 Å.

High-resolution shell	R_{work} (init.)	R_{work} (final)	ΔR_{work}	R_{free} (init.)	R_{final} (final)	ΔR_{free}
2.40 Å → 2.30 Å	0.1782	0.1795	0.0013	0.2316	0.2302	**-0.0014**
2.30 Å → 2.20 Å	0.1837	0.1840	0.0003	0.2341	0.2329	**-0.0012**
2.20 Å → 2.10 Å	0.1875	0.1883	0.0008	0.2365	0.2355	**-0.0010**
2.10 Å → 2.00 Å	0.1913	0.1923	0.0010	0.2376	0.2370	**-0.0006**
2.00 Å → 1.95 Å	0.1962	0.1964	0.0002	0.2402	0.2400	**-0.0002**

The structure contained several problematic regions possessing too weak 2mF_O-DF_C electron density map. Interpretation of these sites was carried out with composite-omit maps (Terwilliger, *et al.*, 2008), feature-enhanced maps (Afonine *et al.*, 2015) and polder maps (Liebschner *et al.*, 2017), calculated using the *Phenix* package (Adams *et al.*, 2010; Liebschner *et al.*, 2019) and NCS-averaged maps (Rossmann and Arnold, 2006) from *Coot*.

The final structure model was refined against all reflections and validated with *Coot* (Emsley *et al.*, 2010), wwPDB validation service (Berman *et al.*, 2003) and *MolProbity* (Williams *et al.*, 2018). The refinement statistics are listed in Table 5. The structure was deposited to the Protein Data Bank (Berman, 2000) under the accession code 8AQ8.

Table 5: Structure refinement statistics for *Sm*TetX.

Resolution (Å)	48.21-1.95 (2.01-1.95)
R_{work}	0.2002 (0.3185)
R_{free} (5 % unique reflections)	0.2410 (0.3289)
R_{all}	0.2022 (0.3175)
CC_{work}	0.964 (0.592)
CC_{free}	0.938 (0.617)
Mean ADP (Å^2)	28.9
No. protein chains in AU	4
No. atoms	11,859
No. water molecules	1,111
Ligands	4x FAD, 6x ethylene glycol, 2x polyethylene glycol, 3x nitrate, 2x sulfate, 4x chloride ion
Matthews coefficient (Å^3 / Da)	2.6
Solvent content (%)	52.6
RMSD bonds from ideal (Å)	0.009
RMSD angles from ideal (°)	1.498
Ramachandran favoured (%)	97
Ramachandran outliers (%)	0
Rama-Z score (Sobolev *et al.*, 2020)	-1.3 ± 0.2

3.1.10 Molecular docking

Molecular docking was performed with *Autodock Vina* (Eberhardt *et al.*, 2021; Trott and Olson, 2009) in the *UCSF Chimera* interface (Pettersen *et al.*, 2004, 2021). A molecule of oxytetracycline was placed in a grid map (25 Å x 24 Å x 20 Å) covering cavities in the *Sm*TetX structure in the vicinity of the isoalloxazine moiety of FAD. The calculated affinity of the most probable binding mode is -8.8 kcal/mol.

3.2 Results

3.2.1 Enzymatic inactivation of oxytetracycline

The inactivation of oxytetracycline (OTC) by *Sm*TetX was tested. The reaction was monitored similarly to previous studies (Forsberg *et al.*, 2015; Gasparrini *et al.*, 2020; Moore *et al.*, 2005; Rudra *et al.*, 2018; Yang *et al.*, 2004) using the decrease in absorption at 400 nm to overcome the overlap of absorption peaks of OTC and nicotinamide adenine dinucleotide phosphate (NADPH) (368 nm and 340 nm, respectively). The observed decrease at 400 nm in absorption spectra, caused by *Sm*TetX, indicates the OTC inactivation (Figure 16abgh). The OTC inactivation occurred also in the absence of NADPH (Figure 16ag). The enzyme also consumes the NADPH cofactor (Figure 16ci).

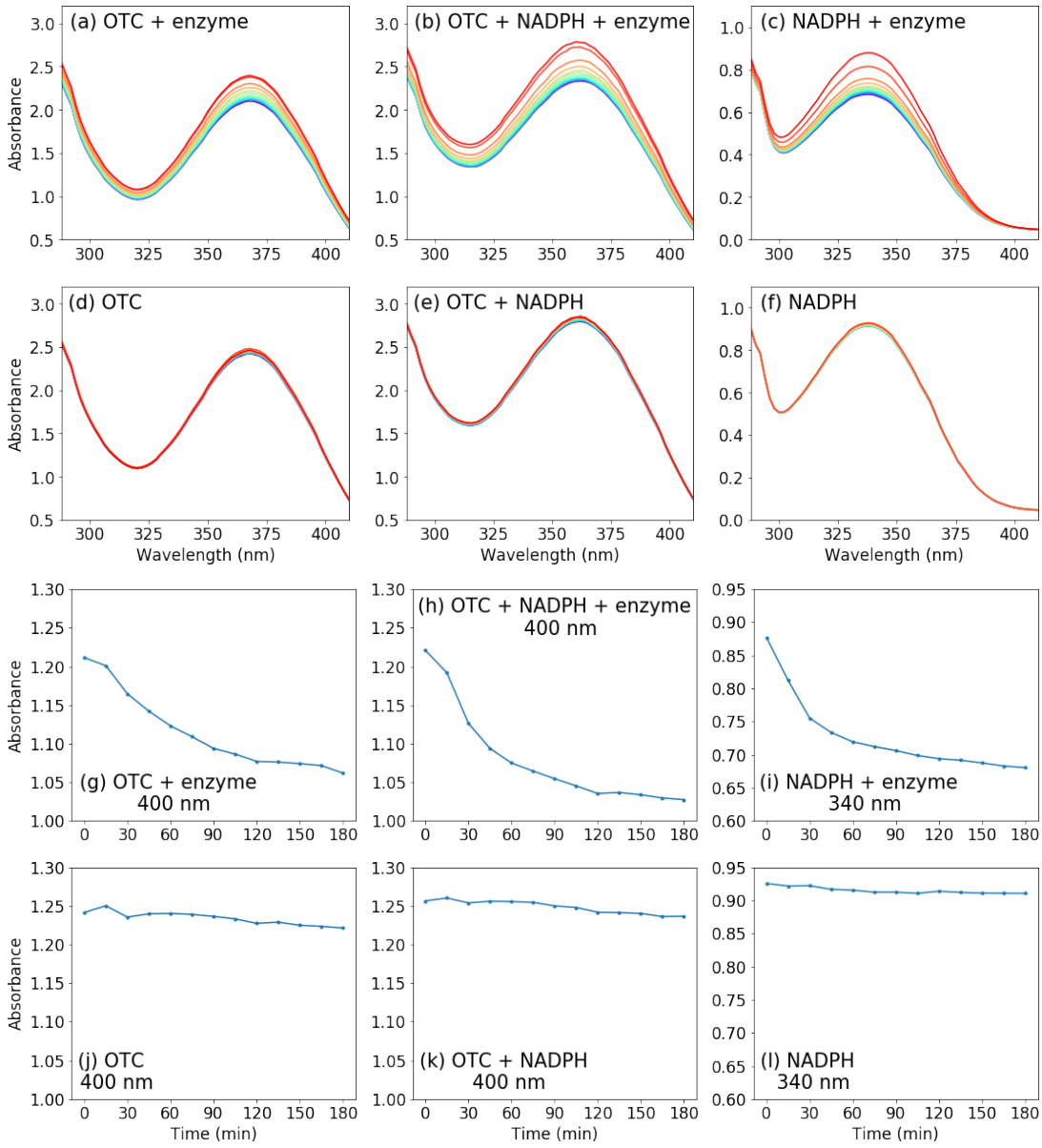


Figure 16: Spectral changes in UV-VIS range of oxytetracycline and NADPH caused by *Sm*TetX. (a-f) Each absorbance was taken at 15-min intervals over the time course of 180 min, plotted in a rainbow colour gradient (from red through green to blue). (g-l) Absorbance decrease related to oxytetracycline and NADPH monitored at 400 nm and 340 nm, respectively.

3.2.2 Overall structure

The crystal structure was solved at 1.95 Å and reveals four protein molecules in the asymmetric unit of the crystal. The pairwise root mean square deviations (RMSD) between the protein chains are lower than 0.2 Å; no significant differences between the chains are observed. A few parts of the protein structure could not be modeled owing to poor quality of electron density, including Met1-Gln2, Gly93-Asp101 and Gly358-Gly364. Similarly, the regions Gln40-Arg42 and Ala135-Asn138 were modeled only for one chain.

The overall fold of *Sm*TetX is similar to other class A FDO structures (Huijbers *et al.*, 2014; Paul *et al.*, 2021; Toplak *et al.*, 2021; van Berkel *et al.*, 2006). The structure is composed of the substrate-binding domain and the FAD-binding domain which are stabilized with one C-terminal α-helix (residues Asp324-Arg364) (Figure 17ab). Between the domains, two large cavities are formed that allow accessing the isoalloxazine moiety of the prosthetic group FAD.

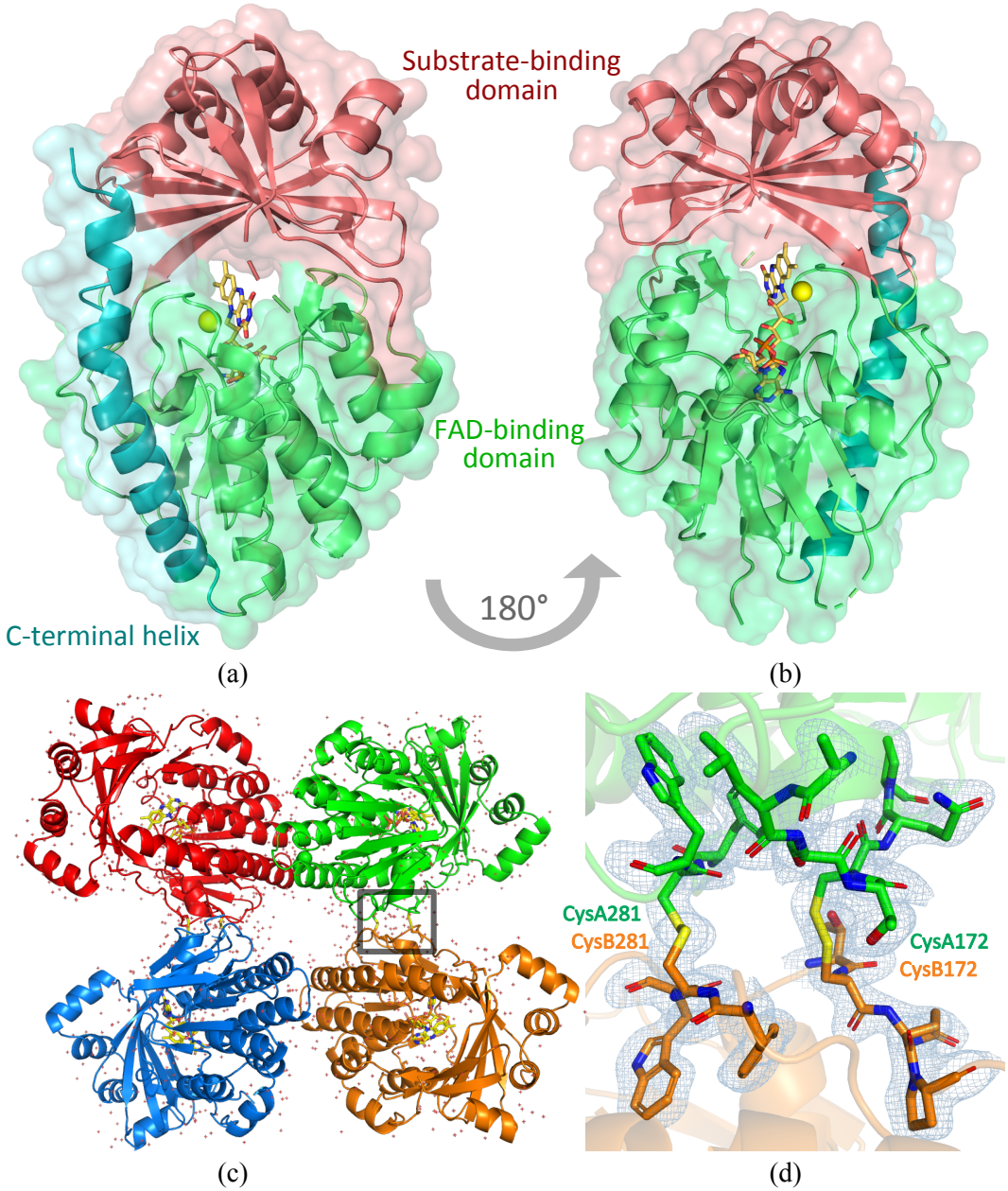


Figure 17: Crystal structure of *Sm*TetX. (a-b) The overall structure of a monomer is divided to the FAD-binding domain in green, the substrate-binding domain in red and the C-terminal helix in blue. Carbon atoms of FAD are displayed in yellow as sticks and a chlorine anion as a yellow sphere. (c) The asymmetric unit is composed of two homodimers that are covalently bound via two cysteine bridges (d), the corresponding $2mF_O$-DF_C electron density map is displayed at 1σ level in blue. The graphics were created in *PyMOL* (Schrödinger and DeLano, 2022).

Results from the small-angle X-ray scattering experiment showed that *Sm*TetX is a monomer in solution (Figure 13). Furthermore, analysis using mass spectrometry confirmed that the protein does not contain any cysteine bridges (Figure 12). A search with *PISA* (Krissinel and Henrick, 2007) did not predict any quaternary structure. However, the four protein chains in the asymmetric unit are assembled into two dimers, covalently linked via two intermolecular disulfide bridges (Cys172(A)-Cys172(B), Cys281(A)-Cys281(B)), see Figure 17cd. The biophysical measurements showed that the dimerization observed in the crystal structure is only a crystallization artifact. On the basis of the high similarity of *Sm*TetX to other class A FDO structures which act as monomers, our structure analysis is not affected.

3.2.3 Composition of domains

The substrate-binding domain (residues Glu75-Ser99 and Pro176-Tyr274) is divided into two main components: four α-helices and a seven-stranded parallel β-sheet (Figure 18).

The FAD-binding domain (residues Met1-His74, Lys100-Leu175 and Ser275-Pro323) is composed of a five-stranded β-sheet (parallel), a three-stranded β-sheet (antiparallel), a two-stranded β-sheet (antiparallel), six α-helices and one 3-10 helix, as assigned with *STRIDE* (Heinig and Frishman, 2004). A βαβ motif of a Rossman fold (Rao and Rossmann, 1973), found at the N-terminus, contains a G*X*G*XX*G sequence motif (residues Gly11-Ala15), providing binding of the FAD adenine moiety (Dym and Eisenberg, 2001; Kleiger and Eisenberg, 2002; Wierenga *et al.*, 1986). The binding of this moiety is further strengthened with an Asparagine-Glycine (DG) motif (residues Asp159-Gly160). Moreover, the FAD pyrophosphate moiety is bound to the protein with a Glycine-Asparagine (GD) motif (Gly292-Asp293) placed in the vicinity of the 3-10 helix (Asp293-Ala295).

The molecule of FAD is well localized in electron density and forms hydrogen bonds with the following residues: Ala15, Glu34, Arg107, Leu129, Asp293 and Val305 (Figure 19ab).

3.2.4 Binding pockets

The prosthetic group FAD can be accessed through two large cavities located between the domains. They will be denoted with respect to the FAD isoalloxazine moiety as *re*-site and *si*-site.

Substrate is most likely bound in the *re*-site (Figure 19c). A large distinct nonpolar region is located, consisting of residues Phe181, Phe203, Phe205, Phe212, Tyr224, Phe351 and Trp354 is located opposite isoalloxazine. On the contrary, residues Asp49, His51, Ser201, Gln214, Cys222 together with the main chain of a turn next to the pyrimidine moiety of FAD isoalloxazine (Pro300, Val301 and Gly302) allow

hydrogen bonding of a substrate. The nitrogen atoms of this turn coordinate a chlorine anion located above the N10 atom of the FAD isoalloxazine ring.

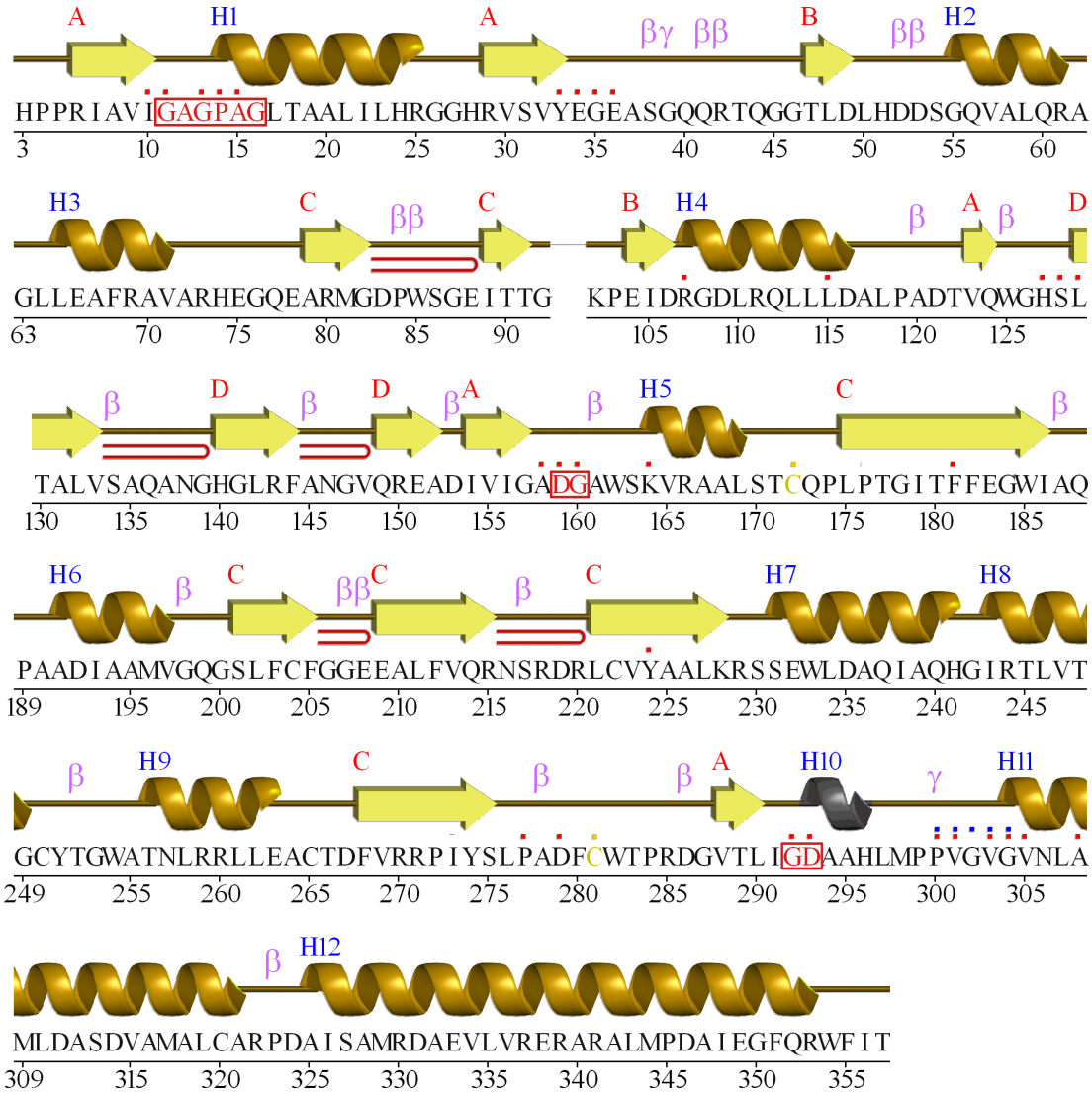


Figure 18: Amino acid sequence of *Sm*TetX with secondary structure elements, structural features and important motifs. Residues interacting with the prosthetic group FAD are marked with red dots and the characteristic motifs G*X*G*XX*G, DG and GD are marked with red rectangles. Residues in contact with the chlorine anion are labeled with blue dots. Cysteine residues that form the intermolecular disulfide bridges are colored in yellow. Helices are labeled in blue and strands by their sheets in red. β-turns and γ-turns are labeled in pink; β-hairpins are marked with red bends. The figure was created with *PDBsum* (Cantelli *et al.*, 2022; Laskowski *et al.*, 2018) and modified.

The whole FAD molecule is accessible from the wide *si*-side cavity and allows binding of the NADPH cofactor, similarly to other enzymes from the class A FMO (Crozier-Reabe and Moran, 2012). The FAD-IN conformation is observed in all four chains of the structure. Nevertheless, the spatial arrangement of this cavity also allows movements to the FAD-OUT conformation, reported *e.g.* in the crystal structure of Tet(50) (Park *et al.*, 2017), with no obvious atomic clashes.

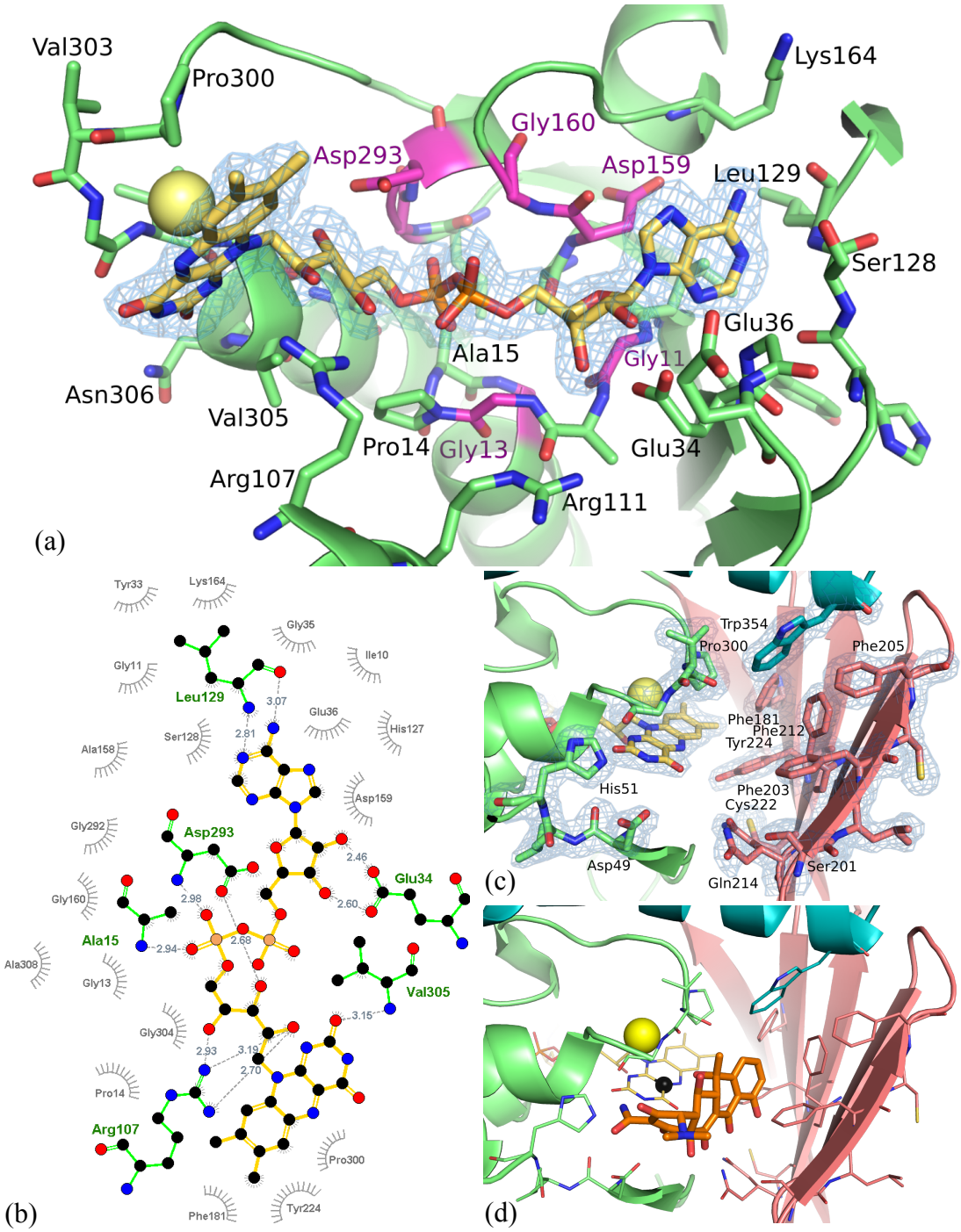


Figure 19: Crystal structure of *Sm*TetX. (a) View from the *si*-site on the FAD binding in combined stick and secondary-structure representation. FAD is colored in yellow. The 2mF_O-DF_C electron density map around FAD is displayed at 1σ level in blue. The residues representing the conserved binding motifs (G*X*G*XX*G, DG and GD) are colored in magenta. (b) Analysis of interactions between FAD and protein residues from the program *LigPlot+* (Laskowski and Swindells, 2011). Hydrogen bridges are displayed in gray with a distance in angstrom. (c) The active site is located in the *re*-site. (d) The orientation of oxytetracycline (orange) in the binding pocket calculated with molecular docking (affinity -8.8 kcal/mol). The C4A atom of FAD is highlighted as a black sphere. The graphics (a) and (c-d) were created in *PyMOL* (Schrödinger and DeLano, 2022).

3.2.5 Molecular docking

In spite of many soaking and cocrystallization experiments, no tetracycline-like ligand was observed in crystals of *Sm*TetX. Thus, possible interactions of the protein with oxytetracycline (OTC) were investigated *in silico*. The molecule of OTC was docked into the *re*-site as well as the *si*-site.

The calculations propose the placement of OTC at the *re*-site, twisted around the pyrimidine moiety of the FAD isoalloxazine ring (Figure 19d). The rather nonpolar ring D of OTC is embedded at the end of the cavity possessing a hydrophobic character owing to residues Phe181, Phe203, Phe212, Tyr224, Phe351 and Trp354. Moreover, OTC can form several hydrogen bonds with the O2, N3 and O4 atoms of the isoalloxazine moiety, with the side chains of Asp49, His51, Gln214 and Asn306, and with the main chain of a turn composed of Pro300, Val301 and Gly302.

Interestingly, the C4a atom of FAD, *i.e.* the key site for modification to the reactive C4a-hydroperoxyflavine, is less than 6 Å distant from several carbon atoms of OTC, including C1, C5a, C6, C6a, C10a, C11, C12.

3.3 Discussion

Activity assays indicate the inactivation of oxytetracycline (OTC) by *Sm*TetX (Figure 16) that propose a classification of the protein to the enzymatic class of tetracycline destructases (Cheng *et al.*, 2021; Markley and Wencewicz, 2018; Yang *et al.*, 2004). However, the catalysis is slower in comparison with the tetracycline descructases (Forsberg *et al.*, 2015). According to the literature, the destructases strictly need the NADPH cofactor for the reduction of FAD prior to the formation of C4a-hydroperoxyflavin that executes an electrophilic attack on a substrate (Crozier-Reabe and Moran, 2012; Toplak *et al.*, 2021). Unexpectedly, a decrease in the OTC absorption spectrum caused by *Sm*TetX took place even without the presence of NADPH (Figure 16ag). This can be explained with a partial presence of the reduced state of FAD of *Sm*TetX. The UV-VIS absorption spectrum of the protein (Figure 10a) supports this hypothesis.

According to our data from multiple experimental methods (SEC chromatography, mass spectrometry – Figure 12, SAXS – Figure 13), *Sm*TetX is in a monomeric state in solution. The solved crystal structure consists of covalently bound dimers (Figure 17cd), which is an artifact of the crystallization method.

The overall fold of *Sm*TetX, composed of two major domains and one C-terminal helix, is characteristic for the class A Flavoprotein Monooxygenases (FMO) (Huijbers *et al.*, 2014; Paul *et al.*, 2021; van Berkel *et al.*, 2006) and is very similar to the tetracycline destructases TetX, TetX2, TetX4 and TetX7 (Cheng *et al.*, 2021; Gasparrini *et al.*, 2020; Volkers *et al.*, 2011), see Figure 21a. The FAD molecule is

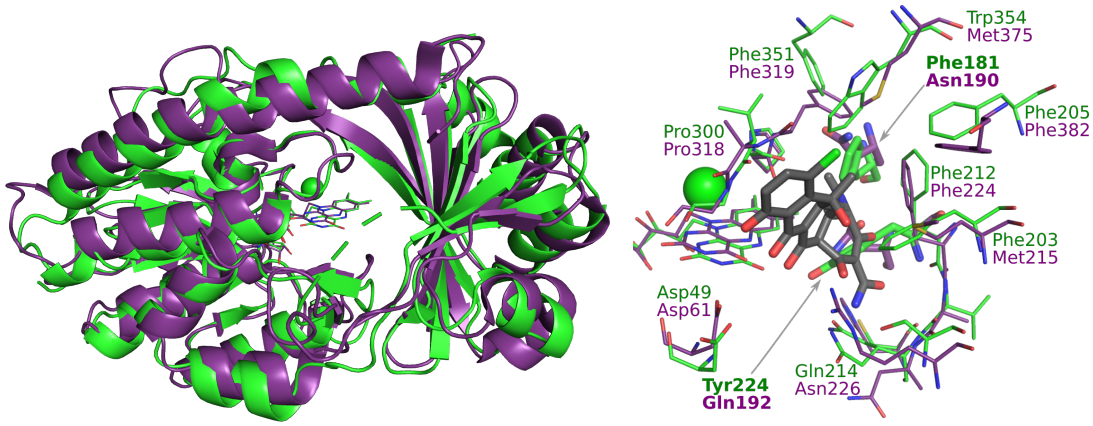
bound non-covalently with structural motifs common for this class (G*X*G*XX*G, DG and GD); however, the tetracycline destructases family lacks the DG motif.

The crucial isoalloxazine moiety of FAD is accessible to the solvent molecules via two cavities. According to molecular docking of oxytetracycline and the structures of complexes of destructases with tetracyclines (Cheng *et al.*, 2021; Park *et al.*, 2017; Volkers *et al.*, 2011, 2013), the substrate-binding pocket is located at the *re*-side cavity (Figure 19c). Despite several similarities, the composition of the substrate-binding pocket is unique in comparison with analogous proteins: the putative reductase AbsH3 from *Streptomyces sp.*, used in molecular replacement, and the tetracycline destructases TetX from Bacteroides *thetaiotaomicron* and Tet(50) (Figure 20). Regions Thr47-His51 and Pro300-Gly302, located in the vicinity of the FAD isoalloxazine moiety, represent the most conserved parts. The *si*-site is supposed to bind the NADPH cofactor, similarly to other class A FDOs.

Figure 20: Multiple structure-based sequence alignment of *Sm*TetX with the putative reductase AbsH3 (Clinger *et al.*, 2021; PDB entry 6N04) and the tetracycline destructases TetX (Volkers *et al.*, 2011; PDB entry 2Y6R) and Tet50 (Park *et al.*, 2017; PDB entry 5TUE). Residues important for substrate binding in the *re*-site of *Sm*TetX are colored in black. Difference, equality or similarity of these residues are marked with red, orange or green background, respectively. *For TetX, TetX2, TetX4 and TetX7, the structure and sequence of these regions are identical (Cheng *et al.*, 2022).

Out of the enzymes with a reported function and structure, *Sm*TetX is the most similar to TetX. The main difference in *Sm*TetX represents residues Phe181 and Tyr224; they expose their nonpolar aromatic ring in contrast to polar Asn190 and Gln192 in TetX, respectively (Figure 21b). Both observed conformations of Gln214 in *Sm*TetX occupy a lot of space in the binding pocket, whereas its analog in TetX (Asn226) is bent out and does not block the site. Nevertheless, the movement of Gln214 to a different conformation, similar to Asn226 in TetX, is possible without any atomic clashes. Moreover, some motifs are shared between *Sm*TetX and TetX, *e.g.* Asp49, His51, Phe212, Pro300 and Gly302 in *Sm*TetX. Residues Phe205 and Phe351 in *Sm*TetX have their phenylalanine analogues in TetX (Phe382 and Phe319), however, they arise from different elements of secondary structure. The aromatic residues Phe203 and Trp354 in *Sm*TetX replace Met203 and Met375 in TetX.

A chloride anion was observed between the pyrimidine moiety of isoalloxazine (*re*-site) and nitrogen atoms of the main chain of the loop Pro300-Val303. Analogous binding of a halogen anion was reported in several structures of class A FMO (Hicks *et al.*, 2013; Matsushita *et al.*, 2020; Nakamoto *et al.*, 2019; Rodríguez Benítez *et al.*, 2019). Nearby this site is supposed to host the oxygen atoms of the reactive hydroperoxy group of C4a-hydroperoxyflavin (Montersino *et al.*, 2013).



(a) (b)

Figure 21: Structure alignment of *Sm*TetX in green with TetX in purple (Volkers *et al.*, 2011; PDB entry 2Y6R), performed in *PyMOL* (Schrödinger and DeLano, 2022), RMSD = 1.53 Å (1375 atoms). (a) The enzymes share the overall fold. (b) Comparison of residues in the substrate binding pocket. The key differences (Phe181 and Tyr224 in *Sm*TetX) are highlighted. The molecule of chlortetracycline from the TetX complex is shown in grey.

The molecular docking confirmed the possibility of OTC binding with the preference into the *re*-site. The calculation proposed a specific ligand orientation, however, this analysis *in silico* is only approximate and can not exactly determine the binding mode.

The low rate of catalytic activity to oxytetracycline together with significant differences in the active site in comparison with TetX indicate that *Sm*TetX could prefer different substrates from tetracyclines. However, the investigation of such more preferred substrates would require conducting very large high-throughput and expensive biochemical experiments that are beyond the scope of this thesis, focused primarily on structure analysis.

Conclusion

Development of methods

The paired refinement is currently the best approach for the determination of the high-resolution cutoff of diffraction data. Our program *PAIREF* makes this protocol automatic and more accessible for the whole community of macromolecular crystallographers. Nowadays, it can be run on multiple platforms (MS Windows, GNU/Linux and Mac) in command-line or graphical user interface using two different refinement engines: *REFMAC5* and *Phenix.refine*. The program is distributed within the *CCP*4 software suite. We placed great emphasis on the use of modern approaches and user-friendliness. Moreover, the automation of the complete cross-validation in paired refinement represents a unique feature implemented in *PAIREF*.

Results using a simulated dataset confirmed that paired refinement produces structures closest to the truth. The application of the routine was demonstrated on several example diffraction datasets in our publications. Paired refinement has become a well understood and widely used method that prevents scientists from discarding reflections that still contain a useful signal. The remaining challenge for future work represents the paired refinement with strongly anisotropic data.

Structure-function analysis of tetracycline-modifying enzyme from *Stenotrophomonas maltophilia*

There are no reports on enzymatic inactivation of tetracycline antibiotics by *Stenotrophomonas maltophilia*. According to the literature, this species relies on other resistance mechanisms in its defense against tetracyclines. Nevertheless, our tetracycline-modifying enzyme from *S. maltophilia* (*Sm*TetX), closely related to the tetracycline destructase TetX, has been reported in genes of several clinical isolates. The experimental work focused on this enzyme revealed several protein features:

- The enzyme is capable of modification of oxytetracycline.
- The overall fold is shared with TetX.
- The structure reveals a novel active site. It is composed of some residues similar to TetX but also of a unique aromatic region (Phe181, Phe203, Phe205, Phe212, Tyr224 and Trp354).
- The protein is a monomer in solution despite the covalent homodimerization observed in the crystal structure.

The obtained results extend the overall knowledge about the broad enzyme class of A FAD-dependent monooxygenases with particular respect to tetracycline destructases. The solved crystal structure of *Sm*TetX can provide the basis for further *in vitro* or *in silico* structural analysis of complexes with ligands – antibiotics or enzyme inhibitors. This could lead to the design of novel inhibitors with potential application in treatment of infections caused by *S. maltophilia* or closely related pathogens.

Involvement in scientific projects

Besides the publications in scientific journals, the author is (or was) also involved in the following projects:

- CAAS – CZ.02.1.01/0.0/0.0/16_019/0000778 – supported by the Ministry of Education, Youth and Sports of the Czech Republic with the European Regional Development Fund,
- CIISB4HEALTH – CZ.02.1.01/0.0/0.0/16_013/0001776 – supported by the Ministry of Education, Youth and Sports of the Czech Republic with the European Regional Development Fund,
- Structure, Function and Application of Biotechnological Enzymes – SGS22/114/OHK4/2T/14 – supported by the Grant Agency of the Czech Technical University in Prague (the author is a proposer of the project),
- Structural Studies of Enzymes for Biotechnologies – SGS19/189/OHK4/3T/14 – supported by the Grant Agency of the Czech Technical University in Prague.

References

Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.-W., Kapral, G.J., Grosse-Kunstleve, R.W., McCoy, A.J., Moriarty, N.W., Oeffner, R., Read, R.J., Richardson, D.C., Richardson, J.S., Terwilliger, T.C. and Zwart, P.H. (2010) '*PHENIX*: a comprehensive Python-based system for macromolecular structure solution', *Acta Crystallographica Section D Biological Crystallography*, 66(2), pp. 213–221. DOI: 10.1107/S0907444909052925.

Adxv (2022) *Adxv - A program to display X-Ray diffraction images*. Available at: https://www.scripps.edu/tainer/arvai/adxv.html (Accessed: 30 June 2022).

Afonine, P.V., Grosse-Kunstleve, R.W., Echols, N., Headd, J.J., Moriarty, N.W., Mustyakimov, M., Terwilliger, T.C., Urzhumtsev, A., Zwart, P.H. and Adams, P.D. (2012) 'Towards automated crystallographic structure refinement with *phenix.refine*', *Acta Crystallographica Section D Biological Crystallography*, 68(4), pp. 352–367. DOI: 10.1107/S0907444912001308.

Afonine, P.V., Moriarty, N.W., Mustyakimov, M., Sobolev, O.V., Terwilliger, T.C., Turk, D., Urzhumtsev, A. and Adams, P.D. (2015) 'FEM: feature-enhanced map', *Acta Crystallographica Section D Biological Crystallography*, 71(3), pp. 646–666. DOI: 10.1107/S1399004714028132.

Alexander, C.G., Wanner, R., Johnson, C.M., Breitsprecher, D., Winter, G., Duhr, S., Baaske, P. and Ferguson, N. (2014) 'Novel microscale approaches for easy, rapid determination of protein stability in academic and commercial settings', *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, 1844(12), pp. 2241–2250. DOI: 10.1016/j.bbapap.2014.09.016.

Arita, M., Karsch-Mizrachi, I. and Cochrane, G. (2021) 'The international nucleotide sequence database collaboration', *Nucleic Acids Research*, 49(D1), pp. D121–D124. DOI: 10.1093/nar/gkaa967.

Assmann, G., Brehm, W. and Diederichs, K. (2016) 'Identification of rogue datasets in serial crystallography', *Journal of Applied Crystallography*, 49(3), pp. 1021–1028. DOI: 10.1107/S1600576716005471.

Battye, T.G.G., Kontogiannis, L., Johnson, O., Powell, H.R. and Leslie, A.G.W. (2011) '*iMOSFLM*: a new graphical interface for diffraction-image processing with *MOSFLM*', *Acta Crystallographica Section D Biological Crystallography*, 67(4), pp. 271–281. DOI: 10.1107/S0907444910048675.

Behling, R. (2015) *Modern diagnostic X-ray sources: technology, manufacturing, reliability*. Boca Raton: CRC Press/Taylor & Francis. ISBN: 978-1-4822-4132-7.

Bergfors, T. (2007) 'Screening and Optimization Methods for Nonautomated Crystallization Laboratories', in J.M. Walker and S. Doublié (eds) *Macromolecular Crystallography Protocols*. Totowa, New Jersey: Humana Press (Methods in Molecular Biology), pp. 131–152. DOI: 10.1007/978-1-59745-209-0_7.

van Berkel, W.J.H., Kamerbeek, N.M. and Fraaije, M.W. (2006) 'Flavoprotein monooxygenases, a diverse class of oxidative biocatalysts', *Journal of Biotechnology*, 124(4), pp. 670–689. DOI: 10.1016/j.jbiotec.2006.03.044.

Berman, H., Henrick, K. and Nakamura, H. (2003) 'Announcing the worldwide Protein Data Bank', *Nature Structural & Molecular Biology*, 10(12), pp. 980–980. DOI: 10.1038/nsb1203-980.

Berman, H.M. (2000) 'The Protein Data Bank', *Nucleic Acids Research*, 28(1), pp. 235–242. DOI: 10.1093/nar/28.1.235.

Berne, B.J. and Pecora, R. (2000) *Dynamic light scattering: with applications to chemistry, biology, and physics*. Dover ed. Mineola, New York: Dover Publications. ISBN: 978-0486411552.

Blanc, E., Roversi, P., Vonrhein, C., Flensburg, C., Lea, S.M. and Bricogne, G. (2004) 'Refinement of severely incomplete structures with maximum likelihood in *BUSTER–TNT*', *Acta Crystallographica Section D Biological Crystallography*, 60(12), pp. 2210–2221. DOI: 10.1107/S0907444904016427.

Boratyn, G.M., Camacho, C., Cooper, P.S., Coulouris, G., Fong, A., Ma, N., Madden, T.L., Matten, W.T., McGinnis, S.D., Merezhuk, Y., Raytselis, Y., Sayers, E.W., Tao, T., Ye, J. and Zaretskaya, I. (2013) 'BLAST: a more efficient report with usability improvements', *Nucleic Acids Research*, 41(W1), pp. W29–W33. DOI: 10.1093/nar/gkt282.

Brodersen, D.E., Clemons, W.M., Carter, A.P., Morgan-Warren, R.J., Wimberly, B.T. and Ramakrishnan, V. (2000) 'The Structural Basis for the Action of the Antibiotics Tetracycline, Pactamycin, and Hygromycin B on the 30S Ribosomal Subunit', *Cell*, 103(7), pp. 1143–1154. DOI: 10.1016/S0092-8674(00)00216-6.

Brönnimann, C. and Trüb, P. (2016) 'Hybrid Pixel Photon Counting X-Ray Detectors for Synchrotron Radiation', in E.J. Jaeschke, S. Khan, J.R. Schneider, and J.B. Hastings (eds) *Synchrotron Light Sources and Free-Electron Lasers*. Cham: Springer International Publishing, pp. 995–1027. DOI: 10.1007/978-3-319-14394-1_36.

Brooke, J.S. (2012) 'Stenotrophomonas maltophilia: an Emerging Global Opportunistic Pathogen', *Clinical Microbiology Reviews*, 25(1), pp. 2–41. DOI: 10.1128/CMR.00019-11.

Brooke, J.S. (2014) 'New strategies against *Stenotrophomonas maltophilia*: a serious worldwide intrinsically drug-resistant opportunistic pathogen', *Expert Review of Anti-infective Therapy*, 12(1), pp. 1–4. DOI: 10.1586/14787210.2014.864553.

Brünger, A.T. (1992) 'Free R value: a novel statistical quantity for assessing the accuracy of crystal structures', *Nature*, 355(6359), pp. 472–475. DOI: 10.1038/355472a0.

Brünger, A.T. (1997) 'Free R value: Cross-validation in crystallography', *Methods in Enzymology*, 277, pp. 366–396. DOI: 10.1016/S0076-6879(97)77021-6.

Cantelli, G., Bateman, A., Brooksbank, C., Petrov, A.I., Malik-Sheriff, R.S., Ide-Smith, M., Hermjakob, H., Flicek, P., Apweiler, R., Birney, E. and McEntyre, J. (2022) 'The European Bioinformatics Institute (EMBL-EBI) in 2021', *Nucleic Acids Research*, 50(D1), pp. D11–D19. DOI: 10.1093/nar/gkab1127.

Casanas, A., Warshamanage, R., Finke, A.D., Panepucci, E., Olieric, V., Nöll, A., Tampé, R., Brandstetter, S., Förster, A., Mueller, M., Schulze-Briese, C., Bunk, O. and Wang, M. (2016) 'EIGER detector: application in macromolecular crystallography', *Acta Crystallographica Section D Structural Biology*, 72(9), pp. 1036–1048. DOI: 10.1107/S2059798316012304.

Chang, Y.-T., Lin, C.-Y., Chen, Y.-H. and Hsueh, P.-R. (2015) 'Update on infections caused by Stenotrophomonas maltophilia with particular attention to resistance mechanisms and therapeutic options', *Frontiers in Microbiology*, 6. DOI: 10.3389/fmicb.2015.00893.

Cheng, Q., Cheung, Y., Liu, C., Chan, E.W.C., Wong, K.Y., Zhang, R. and Chen, S. (2022) 'Functional and phylogenetic analysis of TetX variants to design a new classification system', *Communications Biology*, 5(1), p. 522. DOI: 10.1038/s42003-022-03465-y.

Cheng, Q., Cheung, Y., Liu, C., Xiao, Q., Sun, B., Zhou, J., Chan, E.W.C., Zhang, R. and Chen, S. (2021) 'Structural and mechanistic basis of the high catalytic activity of monooxygenase Tet(X4) on tigecycline', *BMC Biology*, 19(1), p. 262. DOI: 10.1186/s12915-021-01199-7.

Chopra, I. and Roberts, M. (2001) 'Tetracycline Antibiotics: Mode of Action, Applications, Molecular Biology, and Epidemiology of Bacterial Resistance', *Microbiology and Molecular Biology Reviews*, 65(2), pp. 232–260. DOI: 10.1128/MMBR.65.2.232-260.2001.

Clinger, J.A., Wang, X., Cai, W., Zhu, Y., Miller, M.D., Zhan, C., Van Lanen, S.G., Thorson, J.S. and Phillips, G.N. (2021) 'The crystal structure of AbsH3: A putative flavin adenine dinucleotide-dependent reductase in the abyssomicin biosynthesis pathway', *Proteins: Structure, Function, and Bioinformatics*, 89(1), pp. 132–137. DOI: 10.1002/prot.25994.

Crozier-Reabe, K. and Moran, G. (2012) 'Form Follows Function: Structural and Catalytic Variation in the Class A Flavoprotein Monooxygenases', *International Journal of Molecular Sciences*, 13(12), pp. 15601–15639. DOI: 10.3390/ijms131215601.

Cui, C.-Y., He, Q., Jia, Q.-L., Li, C., Chen, C., Wu, X.-T., Zhang, X.-J., Lin, Z.-Y., Zheng, Z.-J., Liao, X.-P., Kreiswirth, B.N., Liu, Y.-H., Chen, L. and Sun, J. (2021) 'Evolutionary Trajectory of the Tet(X) Family: Critical Residue Changes towards High-Level Tigecycline Resistance', *mSystems*. Edited by T.A. Van Laar, 6(3), pp. E00050-21. DOI: 10.1128/mSystems.00050-21.

Diederichs, K. and Karplus, P.A. (1997) 'Improved R-factors for diffraction data analysis in macromolecular crystallography', *Nature Structural Biology*, 4(4), pp. 269–275. DOI: 10.1038/nsb0497-269.

Diederichs, K. and Karplus, P.A. (2013) 'Better models by discarding data?', *Acta Crystallographica Section D Biological Crystallography*, 69(7), pp. 1215–1222. DOI: 10.1107/S0907444913001121.

Dinapoli, R., Bergamaschi, A., Henrich, B., Horisberger, R., Johnson, I., Mozzanica, A., Schmid, E., Schmitt, B., Schreiber, A., Shi, X. and Theidel, G. (2011) 'EIGER: Next generation single photon counting detector for X-ray applications', *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 650(1), pp. 79–83. DOI: 10.1016/j.nima.2010.12.005.

Doublié, S. (ed.) (2007) *Macromolecular crystallography protocols*. Totowa, New Jersey: Humana Press (Methods in molecular biology, v. 363-364). ISBN: 978-1588292926.

Drenth, J. and Mesters, J. (2007) *Principles of Protein X-Ray Crystallography*. 3rd edition. New York: Springer Science+Business Media, LLC. ISBN: 0-387-33334-7.

Ducruix, A. and Giegé, R. (eds) (1992) *Crystallization of nucleic acids and proteins: a practical approach*. Oxford [England]; New York: IRL Press at Oxford University Press (The Practical approach series). DOI: 10.1016/0014-5793(93)80565-C.

Dym, O. and Eisenberg, D. (2001) 'Sequence-structure analysis of FAD-containing proteins', *Protein Science*, 10(9), pp. 1712–1728. DOI: 10.1110/ps.12801.

Eberhardt, J., Santos-Martins, D., Tillack, A.F. and Forli, S. (2021) 'AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings', *Journal of Chemical Information and Modeling*, 61(8), pp. 3891–3898. DOI: 10.1021/acs.jcim.1c00203.

Emsley, P., Lohkamp, B., Scott, W.G. and Cowtan, K. (2010) 'Features and development of *Coot*', *Acta Crystallographica Section D Biological Crystallography*, 66(4), pp. 486–501. DOI: 10.1107/S0907444910007493.

Entsch, B. and Van Berkel, W.J.H. (1995) 'Structure and mechanism of para-hydroxybenzoate hydroxylase', *The FASEB Journal*, 9(7), pp. 476–483. DOI: 10.1096/fasebj.9.7.7737455.

Espes, E., Hansson, B.A.M., Hemberg, O., Otendal, M., Johansson, G., Takman, P. and Tuohimaa, T. (2016) 'Current status of the liquid-metal-jet X-ray source technology and SAXS applications', *Acta Crystallographica Section A Foundations and Advances*, 72(a1), pp. s178–s179. DOI: 10.1107/S2053273316097321.

Esposito, A., Pompilio, A., Bettua, C., Crocetta, V., Giacobazzi, E., Fiscarelli, E., Jousson, O. and Di Bonaventura, G. (2017) 'Evolution of Stenotrophomonas maltophilia in Cystic Fibrosis Lung over Chronic Infection: A Genomic and Phenotypic Population Study', *Frontiers in Microbiology*, 8, p. 1590. DOI: 10.3389/fmicb.2017.01590.

Evans, P. and McCoy, A. (2008) 'An introduction to molecular replacement', *Acta Crystallographica Section D: Biological Crystallography*, 64(1), pp. 1–10. DOI: 10.1107/S0907444907051554.

Evans, P.R. (2011) 'An introduction to data reduction: space-group determination, scaling and intensity statistics', *Acta Crystallographica Section D Biological Crystallography*, 67(4), pp. 282–292. DOI: 10.1107/S090744491003982X.

Evans, P.R. and Murshudov, G.N. (2013) 'How good are my data and what is the resolution?', *Acta Crystallographica Section D Biological Crystallography*, 69(7), pp. 1204–1214. DOI: 10.1107/S0907444913000061.

Fang, L., Chen, C., Cui, C., Li, X., Zhang, Y., Liao, X., Sun, J. and Liu, Y. (2020) 'Emerging High-Level Tigecycline Resistance: Novel Tetracycline Destructases Spread via the Mobile Tet(X)', *BioEssays*, 42(8), p. 2000014. DOI: 10.1002/bies.202000014.

Feigin, L.A. and Svergun, D.I. (1987) *Structure Analysis by Small-Angle X-Ray and Neutron Scattering*. Edited by G.W. Taylor. Boston: Springer US. DOI: 10.1007/978-1-4757-6624-0.

Forsberg, K.J., Patel, S., Wencewicz, T.A. and Dantas, G. (2015) 'The Tetracycline Destructases: A Novel Family of Tetracycline-Inactivating Enzymes', *Chemistry & Biology*, 22(7), pp. 888–897. DOI: 10.1016/j.chembiol.2015.05.017.

Franke, D., Petoukhov, M.V., Konarev, P.V., Panjkovich, A., Tuukkanen, A., Mertens, H.D.T., Kikhney, A.G., Hajizadeh, N.R., Franklin, J.M., Jeffries, C.M. and Svergun, D.I. (2017) 'ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions', *Journal of Applied Crystallography*, 50(4), pp. 1212–1225. DOI: 10.1107/S1600576717007786.

Franke, D. and Svergun, D.I. (2009) '*DAMMIF* , a program for rapid *ab-initio* shape determination in small-angle scattering', *Journal of Applied Crystallography*, 42(2), pp. 342–346. DOI: 10.1107/S0021889809000338.

Gabadinho, J., Beteva, A., Guijarro, M., Rey-Bakaikoa, V., Spruce, D., Bowler, M.W., Brockhauser, S., Flot, D., Gordon, E.J., Hall, D.R., Lavault, B., McCarthy, A.A., McCarthy, J., Mitchell, E., Monaco, S., Mueller-Dieckmann, C., Nurizzo, D., Ravelli, R.B.G., Thibault, X., Walsh, M.A., Leonard, G.A. and McSweeney, S.M. (2010) '*MxCuBE*: a synchrotron beamline control environment customized for macromolecular crystallography experiments', *Journal of Synchrotron Radiation*, 17(5), pp. 700–707. DOI: 10.1107/S0909049510020005.

Gajdács, M. and Urbán, E. (2019) ‘Prevalence and Antibiotic Resistance of *Stenotrophomonas maltophilia* in Respiratory Tract Samples: A 10-Year Epidemiological Snapshot’, *Health Services Research and Managerial Epidemiology*, 6, p. 233339281987077. DOI: 10.1177/2333392819870774.

Garman, E.F. and Schneider, T.R. (1997) ‘Macromolecular Cryocrystallography’, *Journal of Applied Crystallography*, 30(3), pp. 211–237. DOI: 10.1107/S0021889897002677.

Gasparrini, A.J., Markley, J.L., Kumar, H., Wang, B., Fang, L., Irum, S., Symister, C.T., Wallace, M., Burnham, C.-A.D., Andleeb, S., Tolia, N.H., Wencewicz, T.A. and Dantas, G. (2020) ‘Tetracycline-inactivating enzymes from environmental, human commensal, and pathogenic bacteria cause broad-spectrum tetracycline resistance’, *Communications Biology*, 3(1), p. 241. DOI: 10.1038/s42003-020-0966-5.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D. and Bairoch, A. (2005) ‘Protein Identification and Analysis Tools on the ExPASy Server’, in J.M. Walker (ed.) *The Proteomics Protocols Handbook*. Totowa, NJ: Humana Press (Springer Protocols Handbooks), pp. 571–607. DOI: 10.1385/1-59259-890-0:571.

Gil-Gil, T., Martínez, J.L. and Blanco, P. (2020) ‘Mechanisms of antimicrobial resistance in *Stenotrophomonas maltophilia*: a review of current knowledge’, *Expert Review of Anti-infective Therapy*, 18(4), pp. 335–347. DOI: 10.1080/14787210.2020.1730178.

Glady-Croue, J., Niu, X.-Z., Ramsay, J.P., Watkin, E., Murphy, R.J.T. and Croue, J.-P. (2018) ‘Survival of antibiotic resistant bacteria following artificial solar radiation of secondary wastewater effluent’, *Science of The Total Environment*, 626, pp. 1005–1011. DOI: 10.1016/j.scitotenv.2018.01.101.

Glatter, O. (1977) ‘A new method for the evaluation of small-angle scattering data’, *Journal of Applied Crystallography*, 10(5), pp. 415–421. DOI: 10.1107/S0021889877013879.

Glatter, O. and Kratky, O. (eds) (1982) *Small angle x-ray scattering*. London; New York: Academic Press. ISBN: 0-12-286280-5.

Gore, S., Sanz García, E., Hendrickx, P.M.S., Gutmanas, A., Westbrook, J.D., Yang, H., Feng, Z., Baskaran, K., Berrisford, J.M., Hudson, B.P., Ikegawa, Y., Kobayashi, N., Lawson, C.L., Mading, S., Mak, L., Mukhopadhyay, A., Oldfield, T.J., Patwardhan, A., Peisach, E., Sahni, G., Sekharan, M.R., Sen, S., Shao, C., Smart, O.S., Ulrich, E.L., Yamashita, R., Quesada, M., Young, J.Y., Nakamura, H., Markley, J.L., Berman, H.M., Burley, S.K., Velankar, S. and Kleywegt, G.J. (2017) ‘Validation of Structures in the Protein Data Bank’, *Structure*, 25(12), pp. 1916–1927. DOI: 10.1016/j.str.2017.10.009.

Gorrec, F. (2009) ‘The MORPHEUS protein crystallization screen’, *Journal of Applied Crystallography*, 42(6), pp. 1035–1042. DOI: 10.1107/S0021889809042022.

Goujon, M., McWilliam, H., Li, W., Valentin, F., Squizzato, S., Paern, J. and Lopez, R. (2010) ‘A new bioinformatics analysis tools framework at EMBL-EBI’, *Nucleic Acids Research*, 38(Web Server), pp. W695–W699. DOI: 10.1093/nar/gkq313.

Greenfield, N.J. (2006) ‘Using circular dichroism spectra to estimate protein secondary structure’, *Nature Protocols*, 1(6), pp. 2876–2890. DOI: 10.1038/nprot.2006.202.

Groom, C.R., Bruno, I.J., Lightfoot, M.P. and Ward, S.C. (2016) ‘The Cambridge Structural Database’, *Acta Crystallographica Section B Structural Science, Crystal Engineering and Materials*, 72(2), pp. 171–179. DOI: 10.1107/S2052520616003954.

Grossman, T.H. (2016) ‘Tetracycline Antibiotics and Resistance’, *Cold Spring Harbor Perspectives in Medicine*, 6(4), p. a025387. DOI: 10.1101/cshperspect.a025387.

Guinier, A. (1939) ‘La diffraction des rayons X aux très petits angles : application à l’étude de phénomènes ultramicroscopiques’, *Annales de Physique*, 11(12), pp. 161–237. DOI: 10.1051/anphys/193911120161.

Hajizadeh, N.R., Franke, D., Jeffries, C.M. and Svergun, D.I. (2018) ‘Consensus Bayesian assessment of protein molecular mass from solution X-ray scattering data’, *Scientific Reports*, 8(1), p. 7204. DOI: 10.1038/s41598-018-25355-2.

Hållstedt, J., Espes, E. and Adibhatla, A. (2021) ‘MetalJet X-ray source for time-resolved and *in situ* SAXS’, *Acta Crystallographica Section A Foundations and Advances*, 77(a2), pp. C233–C233. DOI: 10.1107/S0108767321094496.

Heinig, M. and Frishman, D. (2004) ‘STRIDE: a web server for secondary structure assignment from known atomic coordinates of proteins’, *Nucleic Acids Research*, 32, pp. W500–W502. DOI: 10.1093/nar/gkh429.

Helliwell, J.R. (2004) *Macromolecular crystallography with synchrotron radiation*. Cambridge; New York: Cambridge University Press. ISBN: 9780521544047.

Helliwell, J.R. (2021) ‘Combining X-rays, neutrons and electrons, and NMR, for precision and accuracy in structure–function studies’, *Acta Crystallographica Section A Foundations and Advances*, 77(3), pp. 173–185. DOI: 10.1107/S205327332100317X.

Helliwell, J.R. (2022) ‘Pre- and Post-publication Verification for Reproducible Data MiningData mining in Macromolecular CrystallographyCrystallography’, in O. Carugo and F. Eisenhaber (eds) *Data Mining Techniques for the Life Sciences*. New York: Springer US (Methods in Molecular Biology), pp. 235–261. DOI: 10.1007/978-1-0716-2095-3_10.

Hemberg, O. (2004) ‘Liquid-metal-jet anode x-ray tube’, *Optical Engineering*, 43(7), p. 1682. DOI: 10.1117/1.1737787.

Hendrickson, W.A., Smith, J.L. and Sheriff, S. (1985) ‘Direct phase determination based on anomalous scattering’, in *Methods in Enzymology*. Elsevier, pp. 41–55. DOI: 10.1016/0076-6879(85)15006-8.

Hicks, K.A., O’Leary, S.E., Begley, T.P. and Ealick, S.E. (2013) ‘Structural and Mechanistic Studies of HpxO, a Novel Flavin Adenine Dinucleotide-Dependent Urate Oxidase from *Klebsiella pneumoniae*’, *Biochemistry*, 52(3), pp. 477–487. DOI: 10.1021/bi301262p.

Huijbers, M.M.E., Montersino, S., Westphal, A.H., Tischler, D. and van Berkel, W.J.H. (2014) ‘Flavin dependent monooxygenases’, *Archives of Biochemistry and Biophysics*, 544, pp. 2–17. DOI: 10.1016/j.abb.2013.12.005.

Hwu, Y. and Margaritondo, G. (2021) ‘Synchrotron radiation and X-ray free-electron lasers (X-FELs) explained to all users, active and potential’, *Journal of Synchrotron Radiation*, 28(3), pp. 1014–1029. DOI: 10.1107/S1600577521003325.

Jenner, L., Starosta, A.L., Terry, D.S., Mikolajka, A., Filonava, L., Yusupov, M., Blanchard, S.C., Wilson, D.N. and Yusupova, G. (2013) ‘Structural basis for potent inhibitory activity of the antibiotic tigecycline during protein synthesis’, *Proceedings of the National Academy of Sciences*, 110(10), pp. 3812–3816. DOI: 10.1073/pnas.1216691110.

Johnson, I., Bergamaschi, A., Billich, H., Cartier, S., Dinapoli, R., Greiffenberg, D., Guizar-Sicairos, M., Henrich, B., Jungmann, J., Mezza, D., Mozzanica, A., Schmitt, B., Shi, X. and Tinti, G. (2014) ‘Eiger: a single-photon counting x-ray detector’, *Journal of Instrumentation*, 9(05), pp. C05032–C05032. DOI: 10.1088/1748-0221/9/05/C05032.

Joosten, R.P., Long, F., Murshudov, G.N. and Perrakis, A. (2014) ‘The *PDB-REDO* server for macromolecular structure model optimization’, *IUCrJ*, 1(4), pp. 213–220. DOI: 10.1107/S2052252514009324.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S.A.A., Ballard, A.J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A.W., Kavukcuoglu, K., Kohli, P. and Hassabis, D. (2021) 'Highly accurate protein structure prediction with AlphaFold', *Nature*, 596(7873), pp. 583–589. DOI: 10.1038/s41586-021-03819-2.

Kabsch, W. (2010a) 'Integration, scaling, space-group assignment and post-refinement', *Acta Crystallographica Section D Biological Crystallography*, 66(2), pp. 133–144. DOI: 10.1107/S0907444909047374.

Kabsch, W. (2010b) '*XDS*', *Acta Crystallographica Section D Biological Crystallography*, 66(2), pp. 125–132. DOI: 10.1107/S0907444909047337.

Karplus, P.A. and Diederichs, K. (2012) 'Linking Crystallographic Model and Data Quality', *Science*, 336(6084), pp. 1030–1033. DOI: 10.1126/science.1218231.

Karplus, P.A. and Diederichs, K. (2015) 'Assessing and maximizing data quality in macromolecular crystallography', *Current Opinion in Structural Biology*, 34, pp. 60–68. DOI: 10.1016/j.sbi.2015.07.003.

Keegan, R.M. and Winn, M.D. (2007) 'Automated search-model discovery and preparation for structure solution by molecular replacement', *Acta Crystallographica Section D Biological Crystallography*, 63(4), pp. 447–457. DOI: 10.1107/S0907444907002661.

Kleiger, G. and Eisenberg, D. (2002) 'GXXXG and GXXXA Motifs Stabilize FAD and NAD(P)-binding Rossmann Folds Through Cα–H···O Hydrogen Bonds and van der Waals Interactions', *Journal of Molecular Biology*, 323(1), pp. 69–76. OI: 10.1016/S0022-2836(02)00885-9.

Koch, E.-E., Brown, G.S. and Moncton, D.E. (eds) (1983) *Handbook on synchrotron radiation*. Amsterdam; New York: North Holland; Elsevier Science Pub. Co. ISBN: 0-444-86709-0.

Kolenko, P., Stránský, J., Koval', T., Malý, M. and Dohnálek, J. (2021) 'SHELIXIR: automation of experimental phasing procedures using SHELXC/D/E', *Journal of Applied Crystallography*, 54(3), pp. 996–1005. DOI: 10.1107/S1600576721002454.

Konarev, P.V. and Svergun, D.I. (2015) '*A posteriori* determination of the useful data range for small-angle scattering experiments on dilute monodisperse systems', *IUCrJ*, 2(3), pp. 352–360. DOI: 10.1107/S2052252515005163.

Konarev, P.V., Volkov, V.V., Sokolova, A.V., Koch, M.H.J. and Svergun, D.I. (2003) '*PRIMUS*: a Windows PC-based system for small-angle scattering data analysis', *Journal of Applied Crystallography*, 36(5), pp. 1277–1282. DOI: 10.1107/S0021889803012779.

Kovalevskiy, O., Nicholls, R.A., Long, F., Carlon, A. and Murshudov, G.N. (2018) 'Overview of refinement procedures within *REFMAC* 5: utilizing data from different sources', *Acta Crystallographica Section D Structural Biology*, 74(3), pp. 215–227. DOI: 10.1107/S2059798318000979.

Kraft, P., Bergamaschi, A., Bronnimann, Ch., Dinapoli, R., Eikenberry, E.F., Graafsma, H., Henrich, B., Johnson, I., Kobas, M., Mozzanica, A., Schleputz, C.M. and Schmitt, B. (2009) 'Characterization and Calibration of PILATUS Detectors', *IEEE Transactions on Nuclear Science*, 56(3), pp. 758–764. DOI: 10.1109/TNS.2008.2009448.

Kraus, I. (1985) *Úvod do strukturní rentgenografie*. Prague: Academia. ISBN: 21-014-85.

Krissinel, E. and Henrick, K. (2007) ‘Inference of Macromolecular Assemblies from Crystalline State’, *Journal of Molecular Biology*, 372(3), pp. 774–797. DOI: 10.1016/j.jmb.2007.05.022.

Krissinel, E., Uski, V., Lebedev, A., Winn, M. and Ballard, C. (2018) ‘Distributed computing for macromolecular crystallography’, *Acta Crystallographica Section D Structural Biology*, 74(2), pp. 143–151. DOI: 10.1107/S2059798317014565.

Krug, M., Weiss, M.S., Heinemann, U. and Mueller, U. (2012) ‘*XDSAPP*: a graphical user interface for the convenient processing of diffraction data using *XDS*’, *Journal of Applied Crystallography*, 45(3), pp. 568–572. DOI: 10.1107/S0021889812011715.

Kwan, T.O.C., Reis, R., Siligardi, G., Hussain, R., Cheruvara, H. and Moraes, I. (2019) ‘Selection of Biophysical Methods for Characterisation of Membrane Proteins’, *International Journal of Molecular Sciences*, 20(10), p. 2605. DOI: 10.3390/ijms20102605.

Laskowski, R.A., Jabłońska, J., Pravda, L., Vařeková, R.S. and Thornton, J.M. (2018) ‘PDBsum: Structural summaries of PDB entries’, *Protein Science*, 27(1), pp. 129–134. DOI: 10.1002/pro.3289.

Laskowski, R.A., MacArthur, M.W., Moss, D.S. and Thornton, J.M. (1993) ‘PROCHECK: a program to check the stereochemical quality of protein structures’, *Journal of Applied Crystallography*, 26(2), pp. 283–291. DOI: 10.1107/S0021889892009944.

Laskowski, R.A. and Swindells, M.B. (2011) ‘LigPlot+: multiple ligand-protein interaction diagrams for drug discovery’, *Journal of Chemical Information and Modeling*, 51(10), pp. 2778–2786. DOI: 10.1021/ci200227u.

Lebedev, A.A. and Isupov, M.N. (2014) ‘Space-group and origin ambiguity in macromolecular structures with pseudo-symmetry and its treatment with the program *Zanuda*’, *Acta Crystallographica Section D Biological Crystallography*, 70(9), pp. 2430–2443. DOI: 10.1107/S1399004714014795.

Leonarski, F., Mozzanica, A., Brückner, M., Lopez-Cuenca, C., Redford, S., Sala, L., Babic, A., Billich, H., Bunk, O., Schmitt, B. and Wang, M. (2020) ‘JUNGFRAU detector for brighter x-ray sources: Solutions for IT and data science challenges in macromolecular crystallography’, *Structural Dynamics*, 7(1), p. 014305. DOI: 10.1063/1.5143480.

Leslie, A.G.W. (2006) ‘The integration of macromolecular diffraction data’, *Acta Crystallographica Section D Biological Crystallography*, 62(1), pp. 48–57. DOI: 10.1107/S0907444905039107.

Lewis, K. (2020) ‘The Science of Antibiotic Discovery’, *Cell*, 181(1), pp. 29–45. DOI: 10.1016/j.cell.2020.02.056.

Li, L. (2005) *Structural Analysis of Cylindrical Particles by Small Angle X-ray Scattering*. Dissertation thesis. University of Bayreuth.

Liebschner, D., Afonine, P.V., Baker, M.L., Bunkóczi, G., Chen, V.B., Croll, T.I., Hintze, B., Hung, L.-W., Jain, S., McCoy, A.J., Moriarty, N.W., Oeffner, R.D., Poon, B.K., Prisant, M.G., Read, R.J., Richardson, J.S., Richardson, D.C., Sammito, M.D., Sobolev, O.V., Stockwell, D.H., Terwilliger, T.C., Urzhumtsev, A.G., Videau, L.L., Williams, C.J. and Adams, P.D. (2019) ‘Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in *Phenix*’, *Acta Crystallographica Section D Structural Biology*, 75(10), pp. 861–877. DOI: 10.1107/S2059798319011471.

Liebschner, D., Afonine, P.V., Moriarty, N.W., Poon, B.K., Sobolev, O.V., Terwilliger, T.C. and Adams, P.D. (2017) ‘Polder maps: improving OMIT maps by excluding bulk solvent’, *Acta Crystallographica Section D Structural Biology*, 73(2), pp. 148–157. DOI: 10.1107/S2059798316018210.

Lira, F., Berg, G. and Martínez, J.L. (2017) 'Double-Face Meets the Bacterial World: The Opportunistic Pathogen Stenotrophomonas maltophilia', *Frontiers in Microbiology*, 8, p. 2190. DOI: 10.3389/fmicb.2017.02190.

Liu, D., Zhai, W., Song, H., Fu, Y., Schwarz, S., He, T., Bai, L., Wang, Y., Walsh, T.R. and Shen, J. (2020) 'Identification of the novel tigecycline resistance gene tet(X6) and its variants in Myroides, Acinetobacter and Proteus of food animal origin', *Journal of Antimicrobial Chemotherapy*, 75(6), pp. 1428–1431. DOI: 10.1093/jac/dkaa037.

Liu, H. and Zwart, P.H. (2012) 'Determining pair distance distribution function from SAXS data using parametric functionals', *Journal of Structural Biology*, 180(1), pp. 226–234. DOI: 10.1016/j.jsb.2012.05.011.

Loeliger, T., Bronnimann, C., Donath, T., Schneebeli, M., Schnyder, R. and Trub, P. (2012) 'The new PILATUS3 ASIC with instant retrigger capability', in *2012 IEEE Nuclear Science Symposium and Medical Imaging Conference Record (NSS/MIC)*. *2012 IEEE Nuclear Science Symposium and Medical Imaging Conference (2012 NSS/MIC)*, Anaheim: IEEE, pp. 610–615. DOI: 10.1109/NSSMIC.2012.6551180.

Long, F., Nicholls, R.A., Emsley, P., Gražulis, S., Merkys, A., Vaitkus, A. and Murshudov, G.N. (2017) '*AceDRG*: a stereochemical description generator for ligands', *Acta Crystallographica Section D Structural Biology*, 73(2), pp. 112–122. DOI: 10.1107/S2059798317000067.

Madeira, F., Park, Y. mi, Lee, J., Buso, N., Gur, T., Madhusoodanan, N., Basutkar, P., Tivey, A.R.N., Potter, S.C., Finn, R.D. and Lopez, R. (2019) 'The EMBL-EBI search and sequence analysis tools APIs in 2019', *Nucleic Acids Research*, 47(W1), pp. W636–W641. DOI: 10.1093/nar/gkz268.

Madeira, F., Pearce, M., Tivey, A.R.N., Basutkar, P., Lee, J., Edbali, O., Madhusoodanan, N., Kolesnikov, A. and Lopez, R. (2022) 'Search and sequence analysis tools services from EMBL-EBI in 2022', *Nucleic Acids Research*, p. gkac240. DOI: 10.1093/nar/gkac240.

Magnusson, A.O., Szekrenyi, A., Joosten, H., Finnigan, J., Charnock, S. and Fessner, W. (2019) 'nanoDSF as screening tool for enzyme libraries and biotechnology development', *The FEBS Journal*, 286(1), pp. 184–204. DOI: 10.1111/febs.14696.

Malý, M. (2018) *Strukturní analýza komplexů oxidázy s ligandem*. Diploma thesis. Czech Technical University in Prague.

Malý, M., Diederichs, K., Dohnálek, J. and Kolenko, P. (2020) 'Paired refinement under the control of *PAIREF*', *IUCrJ*, 7(4), pp. 681–692. DOI: 10.1107/S2052252520005916.

Malý, M., Diederichs, K., Dohnálek, J. and Kolenko, P. (2021) '*PAIREF*: paired refinement also for *Phenix* users', *Acta Crystallographica Section F Structural Biology Communications*, 77(7), pp. 226–229. DOI: 10.1107/S2053230X21006129.

Manalastas-Cantos, K., Konarev, P.V., Hajizadeh, N.R., Kikhney, A.G., Petoukhov, M.V., Molodenskiy, D.S., Panjkovich, A., Mertens, H.D.T., Gruzinov, A., Borges, C., Jeffries, C.M., Svergun, D.I. and Franke, D. (2021) '*ATSAS 3.0*: expanded functionality and new tools for small-angle scattering data analysis', *Journal of Applied Crystallography*, 54(1), pp. 343–355. DOI: 10.1107/S1600576720013412.

Marek, J. and Trávníček, Z. (2002) *Monokrystalová rentgenová strukturní analýza*. Olomouc: Univerzita Palackého, Přírodovědecká fakulta. ISBN: 978-80-244-0551-3.

Margaritondo, G. and Rebernik Ribic, P. (2011) 'A simplified description of X-ray free-electron lasers', *Journal of Synchrotron Radiation*, 18(2), pp. 101–108. DOI: 10.1107/S090904951004896X.

Markley, J.L., Fang, L., Gasparrini, A.J., Symister, C.T., Kumar, H., Tolia, N.H., Dantas, G. and Wencewicz, T.A. (2019) 'Semisynthetic Analogues of Anhydrotetracycline as Inhibitors of Tetracycline Destructase Enzymes', *ACS Infectious Diseases*, 5(4), pp. 618–633. DOI: 10.1021/acsinfecdis.8b00349.

Markley, J.L. and Wencewicz, T.A. (2018) 'Tetracycline-Inactivating Enzymes', *Frontiers in Microbiology*, 9, p. 1058. DOI: 10.3389/fmicb.2018.01058.

Matsushita, T., Kishimoto, S., Hara, K., Hashimoto, H. and Watanabe, K. (2020) 'Structural and Functional Analyses of a Spiro-Carbon-Forming, Highly Promiscuous Epoxidase from Fungal Natural Product Biosynthesis', *Biochemistry*, 59(51), pp. 4787–4792. DOI: 10.1021/acs.biochem.0c00896.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C. and Read, R.J. (2007) '*Phaser* crystallographic software', *Journal of Applied Crystallography*, 40(4), pp. 658–674. DOI: 10.1107/S0021889807021206.

McNicholas, S., Potterton, E., Wilson, K.S. and Noble, M.E.M. (2011) 'Presenting your structures: the *CCP*4*mg* molecular-graphics software', *Acta Crystallographica Section D Biological Crystallography*, 67(4), pp. 386–394. DOI: 10.1107/S0907444911007281.

McPherson, A. (2009) *Introduction to macromolecular crystallography*. 2. ed. Hoboken: Wiley-Blackwell. ISBN: 978-0-470-18590-2.

McWilliam, H., Li, W., Uludag, M., Squizzato, S., Park, Y.M., Buso, N., Cowley, A.P. and Lopez, R. (2013) 'Analysis Tool Web Services from the EMBL-EBI', *Nucleic Acids Research*, 41(W1), pp. W597–W600. DOI: 10.1093/nar/gkt376.

Meyer, A., Dierks, K., Hussein, R., Brillet, K., Brognaro, H. and Betzel, C. (2015) 'Systematic analysis of protein–detergent complexes applying dynamic light scattering to optimize solutions for crystallization trials', *Acta Crystallographica Section F Structural Biology Communications*, 71(1), pp. 75–81. DOI: 10.1107/S2053230X14027149.

Mobilio, S., Boscherini, F. and Meneghini, C. (2015) *Synchrotron Radiation. Basics, methods and applications*. Berlin, Heidelberg: Springer Berlin Heidelberg. DOI: 10.1007/978-3-642-55315-8.

Montersino, S., Orru, R., Barendregt, A., Westphal, A.H., van Duijn, E., Mattevi, A. and van Berkel, W.J.H. (2013) 'Crystal Structure of 3-Hydroxybenzoate 6-Hydroxylase Uncovers Lipid-assisted Flavoprotein Strategy for Regioselective Aromatic Hydroxylation', *Journal of Biological Chemistry*, 288(36), pp. 26235–26245. DOI: 10.1074/jbc.M113.479303.

Moore, I.F., Hughes, D.W. and Wright, G.D. (2005) 'Tigecycline Is Modified by the Flavin-Dependent Monooxygenase TetX', *Biochemistry*, 44(35), pp. 11829–11835. DOI: 10.1021/bi0506066.

Mueller, M., Wang, M. and Schulze-Briese, C. (2012) 'Optimal fine ϕ-slicing for single-photon-counting pixel detectors', *Acta Crystallographica Section D Biological Crystallography*, 68(1), pp. 42–56. DOI: 10.1107/S0907444911049833.

Mueller, U., Förster, R., Hellmig, M., Huschmann, F.U., Kastner, A., Malecki, P., Pühringer, S., Röwer, M., Sparta, K., Steffien, M., Ühlein, M., Wilk, P. and Weiss, M.S. (2015) 'The macromolecular crystallography beamlines at BESSY II of the Helmholtz-Zentrum Berlin: Current status and perspectives', *The European Physical Journal Plus*, 130(7), p. 141. DOI: 10.1140/epjp/i2015-15141-2.

Murray, C.J., Ikuta, K.S., Sharara, F., Swetschinski, L., Robles Aguilar, G., Gray, A., Han, C., Bisignano, C., Rao, P., Wool, E., Johnson, S.C., Browne, A.J., Chipeta, M.G., Fell, F., Hackett, S., Haines-Woodhouse, G., Kashef Hamadani, B.H., Kumaran, E.A.P., McManigal, B., Agarwal, R., Akech, S., Albertson, S., Amuasi, J., Andrews, J., Aravkin, A., Ashley, E., Bailey, F., Baker, S., Basnyat, B., Bekker, A., Bender, R., Bethou, A., Bielicki, J.,

Boonkasidecha, S., Bukosia, J., Carvalheiro, C., Castañeda-Orjuela, C., Chansamouth, V., Chaurasia, S., Chiurchiù, S., Chowdhury, F., Cook, A.J., Cooper, B., Cressey, T.R., Criollo-Mora, E., Cunningham, M., Darboe, S., Day, N.P.J., De Luca, M., Dokova, K., Dramowski, A., Dunachie, S.J., Eckmanns, T., Eibach, D., Emami, A., Feasey, N., Fisher-Pearson, N., Forrest, K., Garrett, D., Gastmeier, P., Giref, A.Z., Greer, R.C., Gupta, V., Haller, S., Haselbeck, A., Hay, S.I., Holm, M., Hopkins, S., Iregbu, K.C., Jacobs, J., Jarovsky, D., Javanmardi, F., Khorana, M., Kissoon, N., Kobeissi, E., Kostyanev, T., Krapp, F., Krumkamp, R., Kumar, A., Kyu, H.H., Lim, C., Limmathurotsakul, D., Loftus, M.J., Lunn, M., Ma, J., Mturi, N., Munera-Huertas, T., Musicha, P., Mussi-Pinhata, M.M., Nakamura, T., Nanavati, R., Nangia, S., Newton, P., Ngoun, C., Novotney, A., Nwakanma, D., Obiero, C.W., Olivas-Martinez, A., Olliaro, P., Ooko, E., Ortiz-Brizuela, E., Peleg, A.Y., Perrone, C., Plakkal, N., Ponce-de-Leon, A., Raad, M., Ramdin, T., Riddell, A., Roberts, T., Robotham, J.V., Roca, A., Rudd, K.E., Russell, N., Schnall, J., Scott, J.A.G., Shivamallappa, M., Sifuentes-Osornio, J., Steenkeste, N., Stewardson, A.J., Stoeva, T., Tasak, N., Thaiprakong, A., Thwaites, G., Turner, C., Turner, P., van Doorn, H.R., Velaphi, S., Vongpradith, A., Vu, H., Walsh, T., Waner, S., Wangrangsimakul, T., Wozniak, T., Zheng, P., Sartorius, B., Lopez, A.D., Stergachis, A., Moore, C., Dolecek, C. and Naghavi, M. (2022) 'Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis', *The Lancet*, 399(10325), pp. 629–655. DOI: 10.1016/S0140-6736(21)02724-0.

Murshudov, G.N., Skubák, P., Lebedev, A.A., Pannu, N.S., Steiner, R.A., Nicholls, R.A., Winn, M.D., Long, F. and Vagin, A.A. (2011) '*REFMAC* 5 for the refinement of macromolecular crystal structures', *Acta Crystallographica Section D Biological Crystallography*, 67(4), pp. 355–367. DOI: 10.1107/S0907444911001314.

Murshudov, G.N., Vagin, A.A. and Dodson, E.J. (1997) 'Refinement of Macromolecular Structures by the Maximum-Likelihood Method', *Acta Crystallographica Section D Biological Crystallography*, 53(3), pp. 240–255. DOI: 10.1107/S0907444996012255.

Mylonas, E. and Svergun, D.I. (2007) 'Accuracy of molecular mass determination of proteins in solution by small-angle X-ray scattering', *Journal of Applied Crystallography*, 40(s1), pp. s245–s249. DOI: 10.1107/S002188980700252X.

Nakamoto, K.D., Perkins, S.W., Campbell, R.G., Bauerle, M.R., Gerwig, T.J., Gerislioglu, S., Wesdemiotis, C., Anderson, M.A., Hicks, K.A. and Snider, M.J. (2019) 'Mechanism of 6-Hydroxynicotinate 3-Monooxygenase, a Flavin-Dependent Decarboxylative Hydroxylase Involved in Bacterial Nicotinic Acid Degradation', *Biochemistry*, 58(13), pp. 1751–1763. DOI: 10.1021/acs.biochem.8b00969.

NCBI Resource Coordinators (2016) 'Database resources of the National Center for Biotechnology Information', *Nucleic Acids Research*, 44(D1), pp. D7–D19. DOI: 10.1093/nar/gkv1290.

Nguyen, F., Starosta, A.L., Arenz, S., Sohmen, D., Dönhöfer, A. and Wilson, D.N. (2014) 'Tetracycline antibiotics and resistance mechanisms', *Biological Chemistry*, 395(5), pp. 559–575. DOI: 10.1515/hsz-2013-0292.

Oeffner, R.D., Afonine, P.V., Millán, C., Sammito, M., Usón, I., Read, R.J. and McCoy, A.J. (2018) 'On the application of the expected log-likelihood gain to decision making in molecular replacement', *Acta Crystallographica Section D Structural Biology*, 74(4), pp. 245–255. DOI: 10.1107/S2059798318004357.

O'Leary, N.A., Wright, M.W., Brister, J.R., Ciufo, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B., Smith-White, B., Ako-Adjei, D., Astashyn, A., Badretdin, A., Bao, Y., Blinkova, O., Brover, V., Chetvernin, V., Choi, J., Cox, E., Ermolaeva, O., Farrell, C.M., Goldfarb, T., Gupta, T., Haft, D., Hatcher, E., Hlavina, W., Joardar, V.S., Kodali, V.K., Li, W., Maglott, D., Masterson, P., McGarvey, K.M., Murphy, M.R., O'Neill, K., Pujar, S., Rangwala, S.H., Rausch, D., Riddick, L.D., Schoch, C., Shkeda, A., Storz, S.S., Sun, H.,

Thibaud-Nissen, F., Tolstoy, I., Tully, R.E., Vatsan, A.R., Wallin, C., Webb, D., Wu, W., Landrum, M.J., Kimchi, A., Tatusova, T., DiCuccio, M., Kitts, P., Murphy, T.D. and Pruitt, K.D. (2016) 'Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation', *Nucleic Acids Research*, 44(D1), pp. D733–D745. DOI: 10.1093/nar/gkv1189.

Orthaber, D., Bergmann, A. and Glatter, O. (2000) 'SAXS experiments on absolute scale with Kratky systems using water as a secondary standard', *Journal of Applied Crystallography*, 33(2), pp. 218–225. DOI: 10.1107/S0021889899015216.

Oscarsson, M., Beteva, A., Flot, D., Gordon, E., Guijarro, M., Leonard, G., McSweeney, S., Monaco, S., Mueller-Dieckmann, C., Nanao, M., Nurizzo, D., Popov, A.N., von Stetten, D., Svensson, O., Rey-Bakaikoa, V., Chado, I., Chavas, L.M.G., Gadea, L., Gourhant, P., Isabet, T., Legrand, P., Savko, M., Sirigu, S., Shepard, W., Thompson, A., Mueller, U., Nan, J., Eguiraun, M., Bolmsten, F., Nardella, A., Milàn-Otero, A., Thunnissen, M., Hellmig, M., Kastner, A., Schmuckermaier, L., Gerlach, M., Feiler, C., Weiss, M.S., Bowler, M.W., Gobbo, A., Papp, G., Sinoir, J., McCarthy, A.A., Karpics, I., Nikolova, M., Bourenkov, G., Schneider, T., Andreu, J., Cuní, G., Juanhuix, J., Boer, R., Fogh, R., Keller, P., Flensburg, C., Paciorek, W., Vonrhein, C., Bricogne, G. and de Sanctis, D. (2019) '*MXCuBE2*: the dawn of *MXCuBE* Collaboration', *Journal of Synchrotron Radiation*, 26(2), pp. 393–405. DOI: 10.1107/S1600577519001267.

Otendal, M., Tuohimaa, T., Vogt, U. and Hertz, H.M. (2008) 'A 9keV electron-impact liquid-gallium-jet x-ray source', *Review of Scientific Instruments*, 79(1), p. 016102. DOI: 10.1063/1.2833838.

Pak, T.R., Altman, D.R., Attie, O., Sebra, R., Hamula, C.L., Lewis, M., Deikus, G., Newman, L.C., Fang, G., Hand, J., Patel, G., Wallach, F., Schadt, E.E., Huprikar, S., van Bakel, H., Kasarskis, A. and Bashir, A. (2015) 'Whole-Genome Sequencing Identifies Emergence of a Quinolone Resistance Mutation in a Case of Stenotrophomonas maltophilia Bacteremia', *Antimicrobial Agents and Chemotherapy*, 59(11), pp. 7117–7120. DOI: 10.1128/AAC.01723-15.

Pantoliano, M.W., Petrella, E.C., Kwasnoski, J.D., Lobanov, V.S., Myslik, J., Graf, E., Carver, T., Asel, E., Springer, B.A., Lane, P. and Salemme, F.R. (2001) 'High-Density Miniaturized Thermal Shift Assays as a General Strategy for Drug Discovery', *SLAS Discovery*, 6(6), pp. 429–440. DOI: 10.1177/108705710100600609.

Park, J., Gasparrini, A.J., Reck, M.R., Symister, C.T., Elliott, J.L., Vogel, J.P., Wencewicz, T.A., Dantas, G. and Tolia, N.H. (2017) 'Plasticity, dynamics, and inhibition of emerging tetracycline resistance enzymes', *Nature Chemical Biology*, 13(7), pp. 730–736. DOI: 10.1038/nchembio.2376.

Paul, C.E., Eggerichs, D., Westphal, A.H., Tischler, D. and van Berkel, W.J.H. (2021) 'Flavoprotein monooxygenases: Versatile biocatalysts', *Biotechnology Advances*, 51, p. 107712. DOI: 10.1016/j.biotechadv.2021.107712.

Pedersen, J.S. (1997) 'Analysis of small-angle scattering data from colloids and polymer solutions: modeling and least-squares fitting', *Advances in Colloid Interface Science*, 70, pp. 171-210. DOI: 10.1016/S0001-8686(97)00312-6

Petoukhov, M.V., Franke, D., Shkumatov, A.V., Tria, G., Kikhney, A.G., Gajda, M., Gorba, C., Mertens, H.D.T., Konarev, P.V. and Svergun, D.I. (2012) 'New developments in the *ATSAS* program package for small-angle scattering data analysis', *Journal of Applied Crystallography*, 45(2), pp. 342–350. DOI: 10.1107/S0021889812007662.

Petoukhov, M.V., Konarev, P.V., Kikhney, A.G. and Svergun, D.I. (2007) '*ATSAS* 2.1 – towards automated and web-supported small-angle scattering data analysis', *Journal of Applied Crystallography*, 40(s1), pp. s223–s228. DOI: 10.1107/S0021889807002853.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C. and Ferrin, T.E. (2004) 'UCSF Chimera – A visualization system for exploratory research and analysis', *Journal of Computational Chemistry*, 25(13), pp. 1605–1612. DOI: 10.1002/jcc.20084.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Meng, E.C., Couch, G.S., Croll, T.I., Morris, J.H. and Ferrin, T.E. (2021) 'UCSF ChimeraX: Structure visualization for researchers, educators, and developers', *Protein Science*, 30(1), pp. 70–82. DOI: 10.1002/pro.3943.

Porod, G. (1951) 'Die Röntgenkleinwinkelstreuung von dichtgepackten kolloiden Systemen', *Kolloid-Zeitschrift*, 124(2), pp. 83–114. DOI: 10.1007/BF01512792.

Puigbò, P., Guzmán, E., Romeu, A. and Garcia-Vallvé, S. (2007) 'OPTIMIZER: a web server for optimizing the codon usage of DNA sequences', *Nucleic Acids Research*, 35, pp. W126–W131. DOI: 10.1093/nar/gkm219.

Ramachandran, G.N., Ramakrishnan, C. and Sasisekharan, V. (1963) 'Stereochemistry of polypeptide chain configurations', *Journal of Molecular Biology*, 7, pp. 95–99. DOI: 10.1016/s0022-2836(63)80023-6.

Rambo, R.P. and Tainer, J.A. (2013) 'Super-Resolution in Solution X-Ray Scattering and Its Applications to Structural Systems Biology', *Annual Review of Biophysics*, 42(1), pp. 415–441. DOI: 10.1146/annurev-biophys-083012-130301.

Rao, S.T. and Rossmann, M.G. (1973) 'Comparison of super-secondary structures in proteins', *Journal of Molecular Biology*, 76(2), pp. 241–256. DOI: 10.1016/0022-2836(73)90388-4.

Roach, D.J., Burton, J.N., Lee, C., Stackhouse, B., Butler-Wu, S.M., Cookson, B.T., Shendure, J. and Salipante, S.J. (2015) 'A Year of Infection in the Intensive Care Unit: Prospective Whole Genome Sequencing of Bacterial Clinical Isolates Reveals Cryptic Transmissions and Novel Microbiota', *PLOS Genetics*. Edited by D. Hughes, 11(7), p. e1005413. DOI: 10.1371/journal.pgen.1005413.

Rodríguez Benítez, A., Tweedy, S.E., Baker Dockrey, S.A., Lukowski, A.L., Wymore, T., Khare, D., Brooks, C.L., Palfey, B.A., Smith, J.L. and Narayan, A.R.H. (2019) 'Structural Basis for Selectivity in Flavin-Dependent Monooxygenase-Catalyzed Oxidative Dearomatization', *ACS Catalysis*, 9(4), pp. 3633–3640. DOI: 10.1021/acscatal.8b04575.

Rossmann, M.G. (1972) *Molecular replacement method*. New York: Gordon and Breach. ISBN: 978-0677139401.

Rossmann, M.G. (ed.) (2001) *International tables for crystallography. F: Crystallography of biological macromolecules*. Dordrecht: Kluwer Academic. ISBN: 978-0-470-66078-2.

Rossmann, M.G. and Arnold, E. (2006) 'Noncrystallographic symmetry averaging of electron density for molecular-replacement phase refinement and extension', in M.G. Rossmann and E. Arnold (eds) *International Tables for Crystallography*. 1st edition. Chester: International Union of Crystallography (International Tables for Crystallography), pp. 279–292. DOI: 10.1107/97809553602060000684.

Rossmann, M.G. and Blow, D.M. (1962) 'The detection of sub-units within the crystallographic asymmetric unit', *Acta Crystallographica*, 15(1), pp. 24–31. DOI: 10.1107/S0365110X62000067.

Roy, A., Kucukural, A. and Zhang, Y. (2010) 'I-TASSER: a unified platform for automated protein structure and function prediction', *Nature Protocols*, 5(4), pp. 725–738. DOI: 10.1038/nprot.2010.5.

Rudra, P., Hurst-Hess, K., Lappierre, P. and Ghosh, P. (2018) 'High Levels of Intrinsic Tetracycline Resistance in Mycobacterium abscessus Are Conferred by a Tetracycline-Modifying Monooxygenase', *Antimicrobial Agents and Chemotherapy*, 62(6), p. e00119. DOI: 10.1128/AAC.00119-18.

Rupp, B. (2009) *Biomolecular Crystallography*. New York: Garland Science. ISBN: 978-0-429-25875-6.

Russo Krauss, I., Merlino, A., Vergara, A. and Sica, F. (2013) 'An Overview of Biological Macromolecule Crystallization', *International Journal of Molecular Sciences*, 14(6), pp. 11643–11691. DOI: 10.3390/ijms140611643.

Sánchez, M.B. (2015) 'Antibiotic resistance in the opportunistic pathogen Stenotrophomonas maltophilia', *Frontiers in Microbiology*, 6. DOI: 10.3389/fmicb.2015.00658.

Schmidt, P.W. (1995) 'Some Fundamental Concepts and Techniques Useful in Small-Angle Scattering Studies of Disordered Solids', in H. Brumberger (ed.) *Modern Aspects of Small-Angle Scattering*. Dordrecht: Springer Netherlands, pp. 1–56. DOI: 10.1007/978-94-015-8457-9_1.

Schnablegger, H. and Singh, Y. (2017) *The SAXS Guide*. 4th edition. Austria: Anton Paar GmbH.

Schreuder, H.A., Prick, P.A.J., Wierenga, R.K., Vriend, G., Wilson, K.S., Hol, W.G.J. and Drenth, J. (1989) 'Crystal structure of the p-hydroxybenzoate hydroxylase-substrate complex refined at 1.9 Å resolution', *Journal of Molecular Biology*, 208(4), pp. 679–696. DOI: 10.1016/0022-2836(89)90158-7.

Schrödinger, LLC and DeLano, W. (2022) 'The PyMOL Molecular Graphics System, Version 2.4'. Available at: http://www.pymol.org/pymol (Accessed: 30 June 2022).

Sheldrick, G.M. (2008) 'A short history of *SHELX*', *Acta Crystallographica Section A Foundations of Crystallography*, 64(1), pp. 112–122. DOI: 10.1107/S0108767307043930.

Sheldrick, G.M. (2010) 'Experimental phasing with *SHELXC* / *D* / *E*: combining chain tracing with density modification', *Acta Crystallographica Section D Biological Crystallography*, 66(4), pp. 479–485. DOI: 10.1107/S0907444909038360.

Sheldrick, G.M. (2015) 'Crystal structure refinement with *SHELXL*', *Acta Crystallographica Section C Structural Chemistry*, 71(1), pp. 3–8. DOI: 10.1107/S2053229614024218.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J.D. and Higgins, D.G. (2011) 'Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega', *Molecular Systems Biology*, 7(1), p. 539. DOI: 10.1038/msb.2011.75.

Smart, O.S., Horský, V., Gore, S., Svobodová Vařeková, R., Bendová, V., Kleywegt, G.J. and Velankar, S. (2018) 'Validation of ligands in macromolecular structures determined by X-ray crystallography', *Acta Crystallographica Section D Structural Biology*, 74(3), pp. 228–236. DOI: 10.1107/S2059798318002541.

Sobolev, O.V., Afonine, P.V., Moriarty, N.W., Hekkelman, M.L., Joosten, R.P., Perrakis, A. and Adams, P.D. (2020) 'A Global Ramachandran Score Identifies Protein Structures with Unlikely Stereochemistry', *Structure*, 28(11), pp. 1249-1258.e2. DOI: 10.1016/j.str.2020.08.005.

Sparta, K.M., Krug, M., Heinemann, U., Mueller, U. and Weiss, M.S. (2016) '*XDSAPP2.0*', *Journal of Applied Crystallography*, 49(3), pp. 1085–1092. DOI: 10.1107/S1600576716004416.

Stetefeld, J., McKenna, S.A. and Patel, T.R. (2016) ‘Dynamic light scattering: a practical guide and applications in biomedical sciences’, *Biophysical Reviews*, 8(4), pp. 409–427. DOI: 10.1007/s12551-016-0218-6.

Stout, G.H. and Jensen, L.H. (1989) *X-ray structure determination: a practical guide*. 2nd edition. New York: Wiley. ISBN: 0-471-60711-8.

Švecová, L., Østergaard, L.H., Skálová, T., Schnorr, K.M., Koval’, T., Kolenko, P., Stránský, J., Sedlák, D., Dušková, J., Trundová, M., Hašek, J. and Dohnálek, J. (2021) ‘Crystallographic fragment screening-based study of a novel FAD-dependent oxidoreductase from *Chaetomium thermophilum*’, *Acta Crystallographica Section D Structural Biology*, 77(6), pp. 755–775. DOI: 10.1107/S2059798321003533.

Svergun, D.I. (1992) ‘Determination of the regularization parameter in indirect-transform methods using perceptual criteria’, *Journal of Applied Crystallography*, 25(4), pp. 495–503. DOI: 10.1107/S0021889892001663.

Svergun, D.I. (1999) ‘Restoring Low Resolution Structure of Biological Macromolecules from Solution Scattering Using Simulated Annealing’, *Biophysical Journal*, 76(6), pp. 2879–2886. DOI: 10.1016/S0006-3495(99)77443-6.

Svergun, D.I., Koch, M.H.J., Timmins, P. and May, R.P. (2013) *Small angle X-ray and neutron scattering from solutions of biological macromolecules*. Oxford: Oxford University Press. ISBN: 978-0-19-963953-3.

Tanaka, S.K., Steenbergen, J. and Villano, S. (2016) ‘Discovery, pharmacology, and clinical profile of omadacycline, a novel aminomethylcycline antibiotic’, *Bioorganic & Medicinal Chemistry*, 24(24), pp. 6409–6419. DOI: 10.1016/j.bmc.2016.07.029.

Terwilliger, T.C., Grosse-Kunstleve, R.W., Afonine, P.V., Moriarty, N.W., Adams, P.D., Read, R.J., Zwart, P.H. and Hung, L.-W. (2008) ‘Iterative-build OMIT maps: map improvement by iterative model building and refinement without model bias’, *Acta Crystallographica Section D Biological Crystallography*, 64(5), pp. 515–524. DOI: 10.1107/S0907444908004319.

Terwilliger, T.C., Grosse-Kunstleve, R.W., Afonine, P.V., Moriarty, N.W., Zwart, P.H., Hung, L.-W., Read, R.J. and Adams, P.D. (2008) ‘Iterative model building, structure refinement and density modification with the *PHENIX AutoBuild* wizard’, *Acta Crystallographica Section D Biological Crystallography*, 64(1), pp. 61–69. DOI: 10.1107/S090744490705024X.

Tickle, I.J., Flensburg, C., Keller, P., Paciorek, W., Sharff, A., Vonrhein, C. and Bricogne, G. (2018) *STARANISO*. Available at: https://staraniso.globalphasing.org/cgi-bin/staraniso.cgi (Accessed: 12 March 2022). Cambridge: Global Phasing Ltd.

Toplak, M., Matthews, A. and Teufel, R. (2021) ‘The devil is in the details: The chemical basis and mechanistic versatility of flavoprotein monooxygenases’, *Archives of Biochemistry and Biophysics*, 698, p. 108732. DOI: 10.1016/j.abb.2020.108732.

Tropea, J.E., Cherry, S. and Waugh, D.S. (2009) ‘Chapter 19: Expression and Purification of Soluble His6-Tagged TEV Protease’, in S. A. Doyle, *High Throughput Protein Expression and Purification: Methods and Protocols*. Totowa: Humana Press (Methods in Molecular Biology), pp. 297–307. DOI: 10.1007/978-1-59745-196-3.

Trott, O. and Olson, A.J. (2009) ‘AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading’, *Journal of Computational Chemistry*, 31(2), pp. 455–461. DOI: 10.1002/jcc.21334.

Usón, I. and Sheldrick, G.M. (2018) ‘An introduction to experimental phasing of macromolecules illustrated by *SHELX*; new autotracing features’, *Acta Crystallographica Section D Structural Biology*, 74(2), pp. 106–116. DOI: 10.1107/S2059798317015121.

Vagin, A. and Lebedev, A. (2015) 'MoRDa, an automatic molecular replacement pipeline', *Acta Crystallographica Section A Foundations and Advances*, 71, p. s19. DOI: 10.1107/S2053273315099672.

Vagin, A. and Teplyakov, A. (1997) '*MOLREP*: an Automated Program for Molecular Replacement', *Journal of Applied Crystallography*, 30(6), pp. 1022–1025. DOI: 10.1107/S0021889897006766.

Valvoda, V., Polcarová, M. and Lukáč, P. (1992) *Základy strukturní analýzy*. Prague: Charles University.

Vincze, T., Posfai, J. and Roberts, R.J. (2003) 'NEBcutter: a program to cleave DNA with restriction enzymes', *Nucleic Acids Research*, 31(13), pp. 3688–3691. DOI: 10.1093/nar/gkg526.

Volkers, G., Damas, J.M., Palm, G.J., Panjikar, S., Soares, C.M. and Hinrichs, W. (2013) 'Putative dioxygen-binding sites and recognition of tigecycline and minocycline in the tetracycline-degrading monooxygenase TetX', *Acta Crystallographica Section D Biological Crystallography*, 69(9), pp. 1758–1767. DOI: 10.1107/S0907444913013802.

Volkers, G., Palm, G.J., Weiss, M.S., Wright, G.D. and Hinrichs, W. (2011) 'Structural basis for a new tetracycline resistance mechanism relying on the TetX monooxygenase', *FEBS Letters*, 585(7), pp. 1061–1066. DOI: 10.1016/j.febslet.2011.03.012.

Volkers, G., Schuldt, L., Palm, G.J., Wright, G.D. and Hinrichs, W. (2010) 'Crystallization and preliminary X-ray crystallographic analysis of the tetracycline-degrading monooxygenase TetX2 from *Bacteroides thetaiotaomicron*', *Acta Crystallographica Section F Structural Biology and Crystallization Communications*, 66(5), pp. 611–614. DOI: 10.1107/S174430911001225X.

Walkiewicz, K., Davlieva, M., Wu, G. and Shamoo, Y. (2011) 'Crystal structure of *Bacteroides thetaiotaomicron* TetX2: A tetracycline degrading monooxygenase at 2.8 Å resolution: Crystal Structure of TetX2', *Proteins: Structure, Function, and Bioinformatics*, 79(7), pp. 2335–2340. DOI: 10.1002/prot.23052.

Wang, L., Liu, D., Lv, Y., Cui, L., Li, Y., Li, T., Song, H., Hao, Y., Shen, J., Wang, Y. and Walsh, T.R. (2019) 'Novel Plasmid-Mediated *tet*(X5) Gene Conferring Resistance to Tigecycline, Eravacycline, and Omadacycline in a Clinical Acinetobacter baumannii Isolate', *Antimicrobial Agents and Chemotherapy*, 64(1), pp. E01326-19. DOI: 10.1128/AAC.01326-19.

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F.T., de Beer, T.A.P., Rempfer, C., Bordoli, L., Lepore, R. and Schwede, T. (2018) 'SWISS-MODEL: homology modelling of protein structures and complexes', *Nucleic Acids Research*, 46(W1), pp. W296–W303. DOI: 10.1093/nar/gky427.

Weeks, C.M., Adams, P.D., Berendzen, J., Brunger, A.T., Dodson, E.J., Grosse-Kunstleve, R.W., Schneider, T.R., Sheldrick, G.M., Terwilliger, T.C., Turkenburg, M.G.W. and Usón, I. (2003) 'Automatic Solution of Heavy-Atom Substructures', in *Methods in Enzymology*. Elsevier, pp. 37–83. DOI: 10.1016/S0076-6879(03)74003-8.

Weiss, M.S. (2001) 'Global indicators of X-ray data quality', *Journal of Applied Crystallography*, 34(2), pp. 130--135. DOI: 10.1107/S0021889800018227.

Whitmore, L. and Wallace, B.A. (2008) 'Protein secondary structure analyses from circular dichroism spectroscopy: Methods and reference databases', *Biopolymers*, 89(5), pp. 392–400. DOI: 10.1002/bip.20853.

Wiedemann, H. (2007) *Particle accelerator physics*. 3rd edition. Berlin Heidelberg: Springer. ISBN: 978-3-540-49043-2.

Wierenga, R.K., Terpstra, P. and Hol, W.G.J. (1986) 'Prediction of the occurrence of the ADP-binding βαβ-fold in proteins, using an amino acid sequence fingerprint', *Journal of Molecular Biology*, 187(1), pp. 101–107. DOI: 10.1016/0022-2836(86)90409-2.

Williams, C.J., Headd, J.J., Moriarty, N.W., Prisant, M.G., Videau, L.L., Deis, L.N., Verma, V., Keedy, D.A., Hintze, B.J., Chen, V.B., Jain, S., Lewis, S.M., Arendall, W.B., Snoeyink, J., Adams, P.D., Lovell, S.C., Richardson, J.S. and Richardson, D.C. (2018) 'MolProbity: More and better reference data for improved all-atom structure validation', *Protein Science*, 27(1), pp. 293–315. DOI: 10.1002/pro.3330.

Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G.W., McCoy, A., McNicholas, S.J., Murshudov, G.N., Pannu, N.S., Potterton, E.A., Powell, H.R., Read, R.J., Vagin, A. and Wilson, K.S. (2011) 'Overview of the *CCP* 4 suite and current developments', *Acta Crystallographica Section D Biological Crystallography*, 67(4), pp. 235–242. DOI: 10.1107/S0907444910045749.

Winter, G., Waterman, D.G., Parkhurst, J.M., Brewster, A.S., Gildea, R.J., Gerstel, M., Fuentes-Montero, L., Vollmar, M., Michels-Clark, T., Young, I.D., Sauter, N.K. and Evans, G. (2018) '*DIALS*: implementation and evaluation of a new integration package', *Acta Crystallographica Section D Structural Biology*, 74(2), pp. 85–97. DOI: 10.1107/S2059798317017235.

Xie, L., Zhou, A., Zhao, J., Tang, Y., Zhao, R., Zhou, Y., Cao, G., Zhong, C. and Li, J. (2021) 'Comparative insights into multiple drug resistance determinants in Stenotrophomonas maltophilia MER1', *Journal of Global Antimicrobial Resistance*, 27, pp. 20–25. DOI: 10.1016/j.jgar.2021.07.017.

Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J. and Zhang, Y. (2015) 'The I-TASSER Suite: protein structure and function prediction', *Nature Methods*, 12(1), pp. 7–8. DOI: 10.1038/nmeth.3213.

Yang, W., Moore, I.F., Koteva, K.P., Bareich, D.C., Hughes, D.W. and Wright, G.D. (2004) 'TetX Is a Flavin-dependent Monooxygenase Conferring Resistance to Tetracycline Antibiotics', *Journal of Biological Chemistry*, 279(50), pp. 52346–52352. DOI: 10.1074/jbc.M409573200.

Zahradník, J., Kolářová, L., Pařízková, H., Kolenko, P. and Schneider, B. (2018) 'Interferons type II and their receptors R1 and R2 in fish species: Evolution, structure, and function', *Fish & Shellfish Immunology*, 79, pp. 140–152. DOI: 10.1016/j.fsi.2018.05.008.

Zhanel, G.G., Esquivel, J., Zelenitsky, S., Lawrence, C.K., Adam, H.J., Golden, A., Hink, R., Berry, L., Schweizer, F., Zhanel, M.A., Bay, D., Lagacé-Wiens, P.R.S., Walkty, A.J., Lynch, J.P. and Karlowsky, J.A. (2020) 'Omadacycline: A Novel Oral and Intravenous Aminomethylcycline Antibiotic Agent', *Drugs*, 80(3), pp. 285–313. DOI: 10.1007/s40265-020-01257-4.

Zhang, Y. (2008) 'I-TASSER server for protein 3D structure prediction', *BMC Bioinformatics*, 9(1), p. 40. DOI: 10.1186/1471-2105-9-40.

List of abbreviations

Bis-Tris	bis-(2-hydroxy-ethyl)-amino-tris(hydroxymethyl)-methane
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
FAD	flavin adenine dinucleotide
FDO	FAD-dependent monooxygenase
NADH	nicotinamide adenine dinucleotide
NADPH	nicotinamide adenine dinucleotide phophate
NCS	non-crystallographic symmetry
Ni-NTA	nickel nitrilotriacetate
OD600	optical density measured at a wavelength of 600 nm
OTC	oxytetracycline
PASEF	parallel accumulation-serial fragmentation
PDB	Protein Data Bank
RMSD	root mean square deviation
SAXS	small-angle X-ray scattering
SEC	size-exclusion chromatography
*Sm*TetX	tetracycline-modifying enzyme from *Stenotrophomonas maltophilia*
TAPS	[tris(hydroxymethyl)methylamino]propanesulfonic acid
TEV	Tobacco Etch Virus
Tris	tris(hydroxymethyl)aminomethane
UHPLC	ultra-high-performance liquid chromatography
UV-VIS	ultra violet-visible spectroscopy

Publications of the author

Articles in peer-reviewed scientific journals

MALÝ, Martin, DIEDERICHS, Kay, DOHNÁLEK, Jan and KOLENKO, Petr, 2020. Paired refinement under the control of *PAIREF*. *IUCrJ*. Vol. 7, no. 4, pp. 681–692. DOI: 10.1107/S2052252520005916.

MALÝ, Martin, DIEDERICHS, Kay, DOHNÁLEK, Jan and KOLENKO, Petr, 2021. *PAIREF*: paired refinement also for *Phenix* users. *Acta Crystallographica Section F Structural Biology Communications*. Vol. 77, no. 7, pp. 226–229. DOI: 10.1107/S2053230X21006129.

KOLENKO, Petr, STRÁNSKÝ, Jan, KOVAĽ, Tomáš, MALÝ, Martin and DOHNÁLEK, Jan, 2021. *SHELIXIR*: automation of experimental phasing procedures using *SHELXC/D/E*. *Journal of Applied Crystallography*. Vol. 54, no. 3, pp. 996–1005. DOI: 10.1107/S1600576721002454.

ADÁMKOVÁ, Kristýna, KOVAĽ, Tomáš, ØSTERGAARD, Lars Henrik, DUŠKOVÁ, Jarmila, MALÝ, Martin, ŠVECOVÁ, Leona, SKÁLOVÁ, Tereza, KOLENKO, Petr and DOHNÁLEK, Jan, 2022. *Acta Crystallographica Section D Structural Biology*. Vol. 78, no. 10., pp. 1194–1209. DOI: 10.1107/S2059798322008397

Conference abstracts of oral presentations

MALÝ, Martin, SKÁLOVÁ, Tereza, ŠVECOVÁ, Leona, DOHNÁLEK, Jan, BLÁHA, Jan, VANĚK, Ondřej, HARLOS, Karl and KOLENKO, Petr. Paired refinement: Impact of reintegration and rescaling. *Proceedings from the 8th Student Scientific Conference on Solid State Physics and Materials*. 2018. Czech Technical University in Prague. ISBN: 978-80-01-06511-2.

MALÝ, Martin, KOLENKO, Petr, SKÁLOVÁ, Tereza, STRÁNSKÝ, Jan, ŠVECOVÁ, Leona and DOHNÁLEK, Jan. Estimation of diffraction limit using paired refinement. *Materials Structure*. 2019. Vol. 26, no. 3, pp. 165–166.

MALÝ, Martin, DIEDERICHS, Kay, DOHNÁLEK, Jan and KOLENKO, Petr. Advantages of complete cross-validation in paired refinement. *Materials Structure*. 2020. Vol. 27, no. 1, p. 5.

MALÝ, Martin, DIEDERICHS, Kay, DOHNÁLEK, Jan, ŠVECOVÁ, Leona and KOLENKO, Petr. Role of FFT grid in paired refinement. *Proceedings from the 9th Student Scientific Conference on Solid State Physics and Materials*. Czech Technical University in Prague. 2020. pp. 83–86. ISBN: 978-80-01-06799-4.

MALÝ, Martin, DIEDERICHS, Kay, DOHNÁLEK, Jan and KOLENKO, Petr. Application of paired refinement in fragment-screening project. *Materials Structure*. 2020. Vol. 27, no. 2, p. 106.

MALÝ, Martin, DIEDERICHS, Kay, DOHNÁLEK, Jan and KOLENKO, Petr. Looking for the optimal resolution cutoff with *PAIREF* GUI. *Book of abstracts of the 1st Student Conference in Structural Biology*. Czech Society for Structural Biology. 2021. p. 16.

MALÝ, Martin, KOLENKO, Petr, PAVLÍČEK, Jiří, STRÁNSKÝ, Jan, ŠVECOVÁ, Leona, ADÁMKOVÁ, Kristýna, HRUBÝ, Jakub, ŠPELDOVÁ, Karolína and DOHNÁLEK, Jan. Automated matrix microseeding improves crystallization of FAD-dependent oxidoreductase. *Proceedings from the 10th Student Scientific Conference on Solid State Physics and Materials*. Czech Technical University in Prague. 2021. pp. 9–14. ISBN: 978-80-01-06931-8.

MALÝ, Martin, DIEDERICHS, Kay, STRÁNSKÝ, Jan, ADÁMKOVÁ, Kristýna, SKÁLOVÁ, Tereza, DOHNÁLEK, Jan and KOLENKO, Petr. Finding the optimal resolution cutoff with *PAIREF*. *Acta Crystallographica Section A Foundations and Advances*. 2021. Vol. 77, part a2, p. C88. DOI: 10.1107/S010876732109591X.

MALÝ, Martin, KOLENKO, Petr, DUŠKOVÁ, Jarmila, KOVAĽ, Tomáš, SKÁLOVÁ, Tereza, TRUNDOVÁ, Mária, STRÁNSKÝ, Jan, ŠVECOVÁ, Leona, ADÁMKOVÁ, Kristýna, HUSŤÁKOVÁ, Blanka and DOHNÁLEK, Jan. Integrative structural analysis of antibiotic-inactivating enzyme from *Stenotrophomonas maltophilia*. *Materials Structure*. 2022. Vol. 28, no. 1, p. 17.

MALÝ, Martin, KOLENKO, Petr, DUŠKOVÁ, Jarmila, KOVAĽ, Tomáš, SKÁLOVÁ, Tereza, TRUNDOVÁ, Mária, STRÁNSKÝ, Jan, ŠVECOVÁ, Leona, ADÁMKOVÁ, Kristýna, HUSŤÁKOVÁ, Blanka and DOHNÁLEK, Jan. Structural insight into antibiotic-inactivating enzyme from *Stenotrophomonas maltophilia*. *Materials Structure*. 2022. Vol. 28, no. 2, pp. 129–130.

MALÝ, Martin, DIEDERICHS, Kay, STRÁNSKÝ, Jan, ADÁMKOVÁ, Kristýna, DOHNÁLEK, Jan and KOLENKO, Petr. Automatic decision on optimal resolution cutoff with *PAIREF*. *Acta Crystallographica Section A Foundations and Advances*. 2022. *The abstract was accepted for publication.*

MALÝ, Martin, KOLENKO, Petr, DUŠKOVÁ, Jarmila, KOVAĽ, Tomáš, SKÁLOVÁ, Tereza, TRUNDOVÁ, Mária, STRÁNSKÝ, Jan, ŠVECOVÁ, Leona, ADÁMKOVÁ, Kristýna, HUSŤÁKOVÁ, Blanka and DOHNÁLEK, Jan. Integrative structural study of antibiotic-inactivating enzyme from *Stenotrophomonas maltophilia*. *Materials Structure*. 2022. Vol. 28, no. 3, p. 185.