Protein structure determination by X-ray crystallography

Running head: Protein X-ray structure determination

Authors: Andrea Ilari, Ph.D. and Carmelinda Savino, Ph.D.

Affiliation: C.N.R.-Istituto di Biologia e Patologia Molecolari c/o Dipartimento di Scienze Biochimiche "A. Rossi Fanelli, Università "La Sapienza", P.le Aldo Moro, 5, 00185 Roma, Italy.

Tel: +39-06-49910910 · Fax:+39-06-4440062

Corrisponding author: Andrea Ilari

Istituto di Biologia e Patologia Molecolari c/o Dipartimento di Scienze Biochimiche "A. Rossi Fanelli, Università "La Sapienza", P.le Aldo Moro, 5, 00185 Roma, Italy.

Tel: +39-06-49910910 · Fax:+39-06-4440062

Email: andrea.ilari@uniroma1.it

Abstract

X-ray biocrystallography is the most powerful method to obtain a macromolecular structure. The improvement of computational technologies in the last years and the development of new and powerful programs to perform calculations together with the enormous increment in the number of protein structures deposited in the Protein Data Bank, render the resolution of new structures easier than in the past. The aim of this chapter is to furnish practical notions useful to solve a new structure. It is impossible to give more than a flavour of what the X-ray crystallographic technique entails in one brief chapter, therefore we focussed our attention on the Molecular Replacement method. Whenever applicable, this method allows the resolution of macromolecular structures starting from a single data set and a search model downloaded from the PDB, with the only aid of computer work.

Key Words

X-ray crystallography; protein crystallization; molecular replacement; coordinates refinement; model building.

1. Introduction

1.1 Protein crystallization

The first requirement for protein structure determination by X-ray crystallography is the attainment of protein crystals diffracting at high resolution. Protein crystallization is mainly a "trial and error" procedure in which the protein is slowly precipitated from its solution. As a general rule, the purer the protein, the better the chances to grow crystals. Growth of protein crystals starts from a super-saturated solution of the macromolecule, and evolves towards a thermodynamically stable state in which the protein is partitioned between a solid phase and the solution. The time required before the equilibrium is reached has a great influence on the final result, which can go from an amorphous or microcrystalline precipitate to large single crystals. The super-saturation conditions can be obtained by addition of precipitating agents (salts, organic solvents and polyethylene glycol polymers) and/or by modifying some of the internal parameters of the solution, like pH, temperature and protein concentration. Since proteins are labile molecules, extreme conditions of precipitation, pH and temperature should be avoided. Protein crystallization involves three main steps (**1**):

1. Determination of protein degree of purity. If the protein is not at least 90-95% pure further

purification will have to be carried out to achieve crystallization .

2. The proteins should be dissolved in a suitable solvent from which it must be precipitated in crystalline form. The solvent is usually a water buffer solution .

3. The solution is brought to super saturation. In this step small aggregates are formed, which are the nuclei for crystal growth. Once nuclei have been formed, actual crystal growth begins.

1.2 Crystal preparation and Data collection

1.2.1 Protein crystals

A crystal is a periodic arrangement of molecules in the three-dimensional space. Molecules precipitating from a solution tend to reach the lowest free energy state. This is often accomplished by packing in a regular way. This regular packing is formed by the repetition in the three space dimensions of the unit cell, defined by three vectors a, b, c and three angles α, β, γ between them. The unit cells contains a number of asymmetric units which coincide with our macromolecule or with more copies of it, related by symmetry operations such as rotations with or without translations. There are 230 different ways to combine the symmetry operations in a crystal, leading to 230 space groups a list of which can be found in the *International Table of Crystallography*, Vol A (**2**). Nevertheless only 65 space group are allowed in protein crystals, because the application of mirror planes and inversion points would change the configuration of amminoacids from L to D, and D-aminoacids are never found in natural protein.

Macromolecule crystals are loosely packed and contain large solvent-filled holes and channels, which normally occupy 40-60% of the crystal volume. For this reason protein crystals are very fragile and have to be handled with care. In order to maintain its water content unaltered, protein crystals should always be kept in their mother liquor or in the saturated vapor of their mother liquor (**3**, **4**). During data collection (see below) X-rays may cause crystal damage due to the formation of free radicals. The best way to avoid damage is crystal freezing. In cryo-crystallography protein crystals are soaked in a solution called “cryo-protectant” where, when frozen, vitrified water, rather then crystalline ice, is formed. In these conditions, crystals exposed to X-rays undergo a negligible radiation damage. Cryo-crystallography usually allows a complete data set to be collected from a single crystal and results in generally higher quality and resolution diffraction

data, while providing more accurate structural information. Normally all measurements, both in house and using synchrotron radiation are performed at 100 K.

1.2.2 X-ray diffraction

X-ray scattering or diffraction is a phenomenon involving both interference and coherent scattering. Two mathematical descriptions of the interference effect were put forward by Max von Laue and W. L. Bragg (**5**, **6**). The simplest description, known as Bragg's law, is presented here. According to Bragg, X-ray diffraction can be viewed as a process similar to reflection by planes of atoms in the crystal. Incident X-rays are scattered by crystal planes, identified by the Miller indices hkl, (see **Note 1**) with an angle of reflection θ. Constructive interference only occurs when the path length difference between rays diffracting from parallel crystal planes is an integral number of wavelengths. When the crystal planes are separated by a distance d, the path length difference is $2d \cdot \sin\theta$. Thus, for constructive interference to occur the following relation must hold true: $n\lambda = 2d \cdot \sin\theta$. As a consequence of the Bragg's law, to 'see' the individual atoms in a structure, the radiation wavelength must be similar to the interatomic distances, typically 0.15 nm (1.5 Å) .

1.2.3 X-ray sources

X-rays are produced in the laboratory by accelerating a beam of electrons emitted by a cathode into an anode, the metal of which dictates what the wavelength of the resulting X-ray will be. Monochromatization is carried out either by using a thin metal foil which absorbs much of the unwanted radiation or by using the intense low-order diffraction from a graphite crystal. To obtain a brighter source, the anode can be made to revolve (rotating anode generator) and is water-cooled to prevent it from melting. An alternative source of X-ray is obtained when a beam of electrons is bent by a magnet. This is the principle behind the synchrotron radiation sources that are capable of producing X-ray beams some thousand times more intense than a rotating anode generator. A consequence of this high-intensity radiation source is that the data collection times have been drastically reduced. A further advantage is that the X-ray spectrum is continuous from around 0.05-0.3 nm (see **Note 2**).

1.2.3 X-ray Detector

In a X-ray diffraction experiment a diffraction pattern is observed that could be regarded as a three-dimensional lattice, reciprocal to the actual crystal lattice (see **Note 3** and **Fig. 1**). For a crystal structure determination the intensities of all diffracted reflections must be measured. To do so, all corresponding reciprocal lattice points must be brought to diffracting conditions by rotating the lattice (*i.e.*, by rotating the crystal) until the required reciprocal lattice points are on a sphere with radius $1/\lambda$. It follows that an X-ray diffraction instrument consists of two parts:

1. a mechanical part for rotating the crystal;
2. a detecting device to measure the position and the intensity of the diffracted reflections.

For a protein structure determination the number of diffracted beams to be recorded is extremely high, of the order of 10^4-10^6 and requires highly efficient hardware. The most efficient and faster devices for data collection in protein crystallography are image plate, and CCD camera (see **Notes 4**, **5**). These instruments are much more sensitive and fast than a X-ray film, reducing considerably the time for exposure and data processing and solving the time-consuming data collection problem.

1.2.4 Data measurement and data processing

Successful data integration depends on the choice of the experimental parameters during data collection. It is therefore crucial that the diffraction experiment is correctly designed and executed. The essence of data collection strategy is to collect every unique reflection at least once. A list of the most important issues that have to be considered is given:

a. The crystal must be single.

b. In order to have a good signal to noise ratio, it is recommended to measure crystal diffraction at the detector edge.

c. The exposure time has to be chosen carefully: it has to be long enough to allow collection of high resolution data (see **Note 6**), but not so long as to cause overload reflections at low resolution and radiation damage.

d. The rotation angle per image should be optimized: too large angle will result in spatial overlap of spots, too small angle will give too many partial spots (see **Note 7**).

e. High data multeplicity will improve the overall quality of the data by reducing random errors and facilitating outlier identification.

Data analysis, performed with the modern data reduction programs, is normally performed in three stages:

1. Autoindexing of one image. The program deduces the lattice type, the crystal unit cell parameters and crystal orientation parameters from a single oscillation image.

2. Indexing of all images. The program compares the diffraction measurements to the spots predicted on the basis of the autoindexing parameters, assignes the hkl indices and calculates the diffraction intensities for each spot in all the collected images.

3. Scaling. The program scales together the data of all the collected images and calculates the structure factor amplitudes for each reflection (identified by the indices hkl).

1.3 Structure determination

The goal of X-ray crystallography is to obtain the distribution of the electron density which is related to the atomic positions in the unit cell, starting from the diffraction data. The electronic density function has the following expression:

$$\rho(x, y, z)=(1/V)\ \Sigma_{hkl}\mathbf{F}_{hkl}\ e^{-2\pi i(hx+ky+lz)} \qquad [\mathbf{1}]$$

where $\mathbf{F}_{hkl}$ are the structure factors, V is the cell volume and h,k,l are the Miller indices. **F** is an imaginary number and can be represented as a vector with a module and a phase. It is possible to easily calculate the amplitude of **F** directly from the X-ray scattering measurements but the information on the phase value would be lost. Different experimental techniques can be used to solve the "phase problem", allowing the building of the protein three-dimensional structure: Multiple Isomorphous Replacement (MIR), Multiple Anomalous Diffraction (MAD) and Molecular Replacement (MR). The last one can be performed by computational calculations using only the native data set.

1.3.1 Molecular Replacement

The Molecular Replacement method consists in fitting a "probe structure" into the experimental unit cell. The probe structure is starting atomic model, from which estimates of the phases can be computed. Such

a model can be the structure of a protein evolutionary related to the unknown one, or even of the same protein from a different crystal form, if available. It is well known that the level of resemblance of two protein structures correlates well with the level of sequence identity (**7**). If the starting model has at least 40% sequence identity with the protein of which the structure is to be determined, the structures are expected to be very similar and molecular replacement will have considerable chances to be successful. These chances progressively decrease with a decrease in structure similarity of the two proteins.

This method is applicable to a large fraction of new structures since the Protein Data Bank (http://www.rscb.org) (**8**) is becoming larger and larger and therefore the probability to find a good model is ever increasing.

Molecular Replacement involves the determination of the orientation and the position of the known structure with respect to the crystallographic axes of the unknown structure, therefore the problem has to be solved in six dimensions. If we call **X** the set of vectors representing the position of the atoms in the probe and **X'** the transformed set, the transformation can be described as:

$$\mathbf{X}' = [R]\,\mathbf{X} + T \qquad [2]$$

Where R represent the rotation matrix, and T the translation vector. In the traditional Molecular Replacement method, Patterson functions (see **Note 8**), calculated for the model and for the experimental data, are compared. This function has the advantage that it can be calculated without phase information. The maps calculated through the two Patterson functions can be superimposed with a good agreement only when the model is correctly oriented and placed in the right position in the unit cell. The calculation of the six variables, defining the orientation and the position of the model, is a computationally expensive problem that requires an enormous amount of calculations to be solved. However, the Patterson function properties allow to divide the problem into two smaller problems: the determination of i) the rotation matrix and ii) the translation vector. This is possible because the Patterson map is a vector map, with peaks corresponding to the positions of vectors between atoms in the unit cell. The Patterson map vectors can be divided into two categories: intra-molecular vectors (self-vectors) and inter-molecular vectors (cross-vectors). Self-vectors (from one atom in the molecule to another atom in the same molecule) depend only on the orientation of the molecule, and not on its position in the cell, therefore they can be exploited in the rotation function. Cross-

vectors depend both on the orientation of the molecule and on its position in the cell, therefore, once the orientation is known, these vectors can be exploited in the translation function.

1.3.2 Rotation function

As mentioned above, the rotation function is based on the observation that the self-vectors depend only on the orientation of the molecule and not on its position in the unit cell. Thus, the rotation matrix can be found by rotating and superposing the model Patterson (calculated as the self-convolution function of the electron density, see **Note 8**) on the observed Patterson (calculated from the experimental intensity). Mathematically, the rotation function can be expressed as a sum of the product of the two Patterson functions at each point :

$$F(R) = \int_r P_{cryst}(u) P_{self}(Ru) du \qquad [3]$$

Where P_{cryst} and P_{self} are the experimental and the calculated Patterson function repectively, R is the rotation matrix and r is the integration radius. In the integration, the volume around the origin where the Patterson map has a large peak is omitted. The radius of integration has a value of the same order of magnitude as the molecule dimensions because the self-vectors are more concentrated near the origin. The programs most frequently used to solve X-ray structures by Molecular Replacement implement the fast rotation function developed by Tony Crowther, who realised that the rotation function can be computed more quickly using the Fast Fourier Transform, expressing the Patterson maps as spherical harmonics (**9**).

1.3.3 Translation function

Once the orientation matrix of the molecule in the experimental cell is found, the next step is the determination of the translation vector. This operation is equivalent to find the absolute position of the molecule. When the molecule, assuming it is correctly rotate in the cell, is translated all the intermolecular vectors change**.** Therefore, only when the molecules in the crystal are in the correct position, the cross-vectors Patterson functions calculated from the observed data and from the model superimpose with a good agreement. The translation function can be described as:

$$T(t) = \int_v P_{cryst}(u) P_{cross}(ut) du \qquad [4]$$

Where P_{cryst} is the experimental Patterson function, whereas P_{cross} is the Patterson function calculated

from the probe oriented in the experimental crystal, t is the translation vector and u is the intermolecular vector between two symmetry-related molecule.

1.4 Structure refinement

Once the phase has been determined for example with the molecular replacement method, an electron density map can be calculated and interpreted in terms of the polypeptide chain. If the major part of the model backbone can be fitted successfully in the electronic density map, the structure refinement can begin. Refinement is performed adjusting the model in order to find a closer agreement between the calculated and the observed structure factors. The adjustment of the model consists in changing the three positional parameters (x, y, z) and the isotropic temperature factors B (see **Note 9**) for all the atoms in the structure, except the hydrogen atoms. The refinement techniques in protein X-ray crystallography are based on the least squares minimization and depend greatly on the ratio of the number of independent observations to variable parameters. Since the protein crystals diffract very weakly, the errors in the data are often very high and more than 5 intensity measurements for each parameter are necessary to refine protein structures. Generally, the problem is poorly over-determined (the ratio is around 2) or sometimes under-determinated with the ratio below 1.0. Different methods are availeble to solve this problem. One of the most commonly used is the Stereochemically Restrained Least Squares Refinement, which increases the number of the observations by adding stereo-chemical restraints (**10**). The function to minimize consists in a crystallographic term and several stereochemical terms:

$$Q = \Sigma\, w_{hkl}\, \{|F_{obs}|-|F_{cal}|\}^2 + \Sigma\, w_D\, (d_{ideal}-d_{model})^2 + \Sigma\, w_T\, (X_{ideal}-X_{model})^2 + \Sigma\, w_P\, (P_{ideal}-P_{model})^2 +$$

$$+ \Sigma\, w_{NB}(E_{min}-E_{model})^2 + \Sigma\, w_C\, (V_{ideal}-V_{model})^2 \qquad [5]$$

Where w terms indicate weighting parameters: "w_{hkl}" is the usual X-ray restraint, "w_D" restrains the distance (d) between atoms thus defining bond length, bond angles and dihedral angles, "w_T" restrains torsion angles (X), "w_P" imposes the planarity of the aromatic rings (P), "w_{NB}" introduces restraints for non bonded and Van der Waals contacts (E), and finally "w_C" restrains the configuration to the correct enantiomer (V). The crystallographic term is calculated from the difference between the experimental structure factor amplitudes F_{obs} and the structure factor amplitudes calculated from the model F_{calc}. The stereochemical terms are calculated as the difference between the values calculated from the model and the

corresponding ideal values. The ideal values for the geometrical parameters are those measured for small molecules and peptides. The refinement program minimizes the overall function by calculating the shifts in coordinates that will give its minimum value by least squares fitting method. The classical least square method can produce over-fitting artefacts by moving faster towards agreement with structure factor amplitudes than towards correctness of the phases, because its shift directions assume the current model phases to be error-free constants. The refinement programs, developed more recently, use the Maximum Likelihood method that allows a representation of the uncertainty on the phases, so that the latter can be used with more caution (see **Note 10**).

Another popular refinement method is known as "simulated annealing" in which an energy function that combines the X-ray term with a potential energy function comprising terms for bond stretching, bond angle bending, torsion potentials and van der Waals interactions, is minimized (**11**).

The parameter used for estimating the correctness of a model in the refinement process is the crystallographic R factor (R_{cryst}) that is usually the sum of the absolute difference between observed and calculated over the sum of observed structure factor amplitudes:

$$R_{cryst} = \Sigma| F_{obs} - F_{calc}| / \Sigma |Fobs| \qquad [6]$$

Use the R_{cryst} as a guide in the refinement process could be dangerous because it often leads to over-refine the model. For this reason, it is recommended to use also the so-called R_{free} parameter which is similar to R_{cryst}, except for the fact that it is calculated from a fraction of the collected data that has been randomly choosen to be excluded from refinement and maps calculation. In this way, the R_{free} calculation is independent from the refinement process and "phase bias" is not introduced.

During the refinement process both R factors should decrease reaching a value in the 10 to 20 % range.

1.4.1 Model building

A key stage in the crystallographic investigation of an unknown structure is the creation of an atomic model. In macromolecular crystallography, the resolution of experimentally phased maps is rarely high enough so that the atoms are visible. However, the development of modern data collection techniques (cryo-crystallography, synchrotron sources) has determined remarkable improvement in the map quality, which, in turn, has made atomic model building easier. Two types of maps are used to build the model: the "2Fo-Fc"

map and the “Fo-Fc” map. The first one is used to build the protein model backbone and is obtained by substituting the term |2Fo-Fc| exp(-iφcalc) to the structure factor term in the equation of the electronic density ([**1**])**.** The “Fo-Fc” map helps the biocrystallographer to build difficult parts of the model and to find the correct conformation for the side chains; moreover, it is used to add solvent and ligand molecules to the model. The “Fo-Fc” map is obtained by substituting the term |Fo-Fc| exp(-iφcalc) to the structure factor term in the equation of the electronic density.

2.Materials

2.1 Crystallization and crystal preparation

1. Hampton Research crystallization and cryoprotection kits.
2. Protein more than 95 % pure, at a concentration between 5 and 15 mg/ml.
3. VDX plates to set up crystallization trials by hanging drop method.
4. Siliconized glass cover slides and vacuum grease.
5. Magnetic crystal caps, and mounted cryoloops.
6. Cryo tong and crystal wand.
7. Dewars to conserve and transport crystals at nitrogen liquid temperature.

2.2 Data measurement and Data processing

1. Goniometer head.
2. HKL suite (XDisplay, Denzo and Scalepack) for Macromolecular Crystallography.

2.3 Molecular Replacement

1. Linux boxes or Silicon Graphics computers
2. One of the following programs: AMoRe (freely available), MolRep (freely available), Xplor, CNS

2.4 Refinement and Model Building

1. Collaborative computational Project n°4 interactive (CCP4i) suite containing programs to manipulate the data sets, solve the structure and refine the model and calculate the maps (freely available)**.**
2. One of the following programs: QUANTA (Molecular Structure Inc.), Xfit , Coot, O to build the model. The

last three are freely available

3. Refmac5 (CCP4i package), Xplor, CNS, programs to refine the model

3. Methods

3.1 Crystallization and crystal preparation

Precise rules to obtain suitable single protein crystals have not been defined yet. For this reason, protein crystallization is mostly a trial and error procedure. This can summarized in three steps:

1. check protein sample purity, which has to be around 90-95 %;

2. slow increase the precipitating agent concentration (PEGs, salts or organic solvents) in order to favour protein aggregation;

3. change of pH and/or temperature.

It is usually necessary to carry out a great number of experiments to determine the best crystallization conditions whilst using a minimum amount of protein per experiment. The protein concentration should be about 10mg/ml, therefore 1mg of purified protein is sufficient to perform about 100 crystallization experiments. Crystallization can be carried out by using different techniques, the most used of which are: liquid-liquid diffusion methods, crystallization under dialysis and vapour diffusion technique. The latter is described in detail since it is easy to set up and allows the biocrystallographer to utilize a minimum protein amount. The vapour diffusion technique can performed in two ways: the "hanging drop" and the "sitting drop" methods.

1. In the "hanging drop" method, drops are prepared on a siliconized microscope glass cover slip by mixing 1-5 μl of protein solution with the same volume of precipitant solution. The slip is placed upside-down over a depression in a tray; the depression is partly filled (about 1ml) with the required precipitant solution (reservoir solution). The chamber is sealed by applying grease to the circumference of depression before the cover slip is put into place (**Figure 2a**)

2. The "sitting drop" method is preferable when the protein solution has a low surface tension and the equilibration rate between drop solution and reservoir solution needs to be slowed down. A schematic diagram of a sitting drop vessel is shown in **Figure 2b**.

The parameters that can be varied include: nature and concentration of the precipitating agent; buffers to explore the entire pH range; additional salts and detergents; etc.

3.2 Crystal cryoprotection

The most widely used cryo-mounting method consists of the suspension of the crystal in a film of a "antifreeze" solution, held by surface tension across a small diameter loop of fiber and the quick insertion of it into a gaseous nitrogen stream. The cryo-protected solution is obtained by adding, cryo protectant agents such as glycerol, ethylene glycol, MPD (2-Methyl-2,4-pentandiol)or low molecular weight PEG (polyethylene glycol) to the precipitant solution. The crystal is immerged in this solution for a few seconds prior to being flash-frozen. The method places little mechanical stress on the crystal, so it is excellent for fragile samples. Loops are made from very fine (~10 μm diameter) fibers of nylon. As some crystals degrade in growth and harvest solutions, liquid nitrogen storage is an excellent way to stabilize crystals for long periods (**12**). This system is particularly useful when preparing samples for data collection at synchrotron radiation sources, in that, by minimizing the time required by sample preparation, it allows to use the limited time available at these facilities to collect data.

3.3 Data measurement

Once the crystal is placed in the fiber loop, the latter must be attached to a goniometer head. This gadget has two perpendicular arcs that allow rotation of the crystal along two perpendicular axes. Additionally, its upper part can be moved along two perpendicular sledges for further adjustment and centering of the crystal. The goniometer head must be screwed on to a detector, making sure that the crystal is in the X-ray beam. In agreement with the Bragg's law, the crystal-to-detector distance should be as low as possible to obtain the maximum resolution together with a good separation between diffraction spots. Generally a distance of 150 mm allows to collect high quality data sets with a good resolution (i.e. lower than 2.0 Å) for protein crystals with unit cell dimensions around 60-80 Å**.** Long unit cell dimensions (a, b and/or c longer than 150 Å) , large mosaicity (see **Notes 11**) (more than 1.0 degree) or large oscillation range (more than 1.0 degree), are all factor affecting spot separations and causing potential reflection overlaps.

Data collection is best performed interactively, with immediate data processing to get a fast feedback

during data collection. This strategy permits to avoid gross inefficiencies in the setup of the experiment, consisting, for example, in incomplete data sets and/or reflection overlaps and/or large percentages of overloaded reflections.

3.4 Data processing

The basic principles involved in the procedure for integrating diffraction data from macromolecules are common to many data integration programs currently in use. There we describe the data processing performed by the HKL2000 suite (**13**). The currently used data processing methods exploit automated subroutines for indexing the X-ray crystal data collection, which means assigning the correct hkl index to each spot on a diffraction image (**Figure 3**).

1. Peak search. The first automatic step is the peak search, which chooses the most intense spots to be used by the autoindexing subroutine. Peaks are measured in a single oscillation image, which for protein crystals, requires 0.2-1.0 oscillation degrees.

2. Autoindexing of one image. If autoindexing succeeds, a good match between the observed diffraction pattern and predictions is obtained. The auto-indexing permits the identification of the space group and the determination of the cell parameters (see **Note 12** and **Table I**). Indeed others parameters have to be refined. Among them the most important are the crystal and detector orientation parameters, the centre of the direct beam and the crystal-to-detector distance.

3. Autoindexing of all the images. The autoindexing procedure, together with refinement, is repeated for all diffraction images.

Data are processed using a component program of the HKL2000 suite called Denzo. The scaling and merging of indexed data, as well as the global refinement of crystal parameters, is performed with the program Scalepack which is another HKL2000 suite component. The values of unit cell parameters refined from a single image may be quite imprecise. Therefore a post-refinement procedure is implemented in the program to allow for separate refinements of the orientation of each image while using the same unit cell for the whole data set. The quality of X-ray data is firstly assessed by statistical parameters reported in the scale.log file. The first important parameter is the I/σ (I: intensity of the signal, and σ the standard deviation), *i.e*. the signal to noise ratio, which is also used to estimate the maximum resolution. The second parameter is

χ^2 which is closely related to I/σ (see **Note 13**), which the program tries to bring close to 1.0 by manipulating the error model. Another important parameter is the R_{sym}, which is a disagreement index between symmetry related reflections and of which the average value should be below 10% (see **Note 14**). The output of the data processing procedure is a file with suffix .hkl, containing all the measured intensities with their relative σ values and the corresponding hkl indices. Using the program Truncate implemented in the CCP4i suite (**14**), it is possible to calculate: the structure factor amplitudes from the intensities by the French and Wilson method (**15**), the Wilson plot to estimate an overall B factor (see **Note 9**) and an absolute scale factor, and intensity statistics to evaluate the correctness of the data reduction procedure. The truncate output data are stored in a file that usually has the extension .mtz.

3.5 Molecular Replacement

1. Search model. The first operation is searching the Data bases for probe structure similar to the structure to be solved. Since we do not know the structural identity of our protein with the homologous ones we use as guide the sequence identity. Proteins showing high degree of sequence similarity with our "query" protein can be identified in protein sequence data bases using sequence comparison methods such as BLAST (**16**) (see chapter on protein structure prediction for details). The protein of known three-dimensional structure showing the highest sequence identity with our query protein is generally used as search model.

2. Files preparation. The Pdb file of the search probe has to be downloaded from the Protein Data Bank. The file has to be manipulated before the Molecular Replacement is performed. The water molecules as well as the ligand molecules have to be removed from the file. The structure can be transformed in a polyalanine search probe to avoid model bias during the Molecular Replacement procedure (see **Note 15**). The other file we need to perform the molecular replacement is the file .mtz resulting from the data processing (see point *3.4*), containing information about crystal space group, cell dimensions, molecules per unit cell and a list of the collected experimental reflections.

3. Molecular Replacement. The Molecular Replacement procedure consists in Rotation and Translation searches to put the probe structure in the correct position in the experimental cell. This operation can be done using different programs, the most used of which is AMoRe (**17**). In this chapter, we describe briefly the use of MolRep (**18**) one of the most recent program to solve macromolecular structures by Molecular

Replacement. This program belongs to the CCP4i suite and is automated and user friendly. The program performs sequentially rotation followed by translation searches. The only input files to upload are the .mtz and the .pdb files. The values of two parameters have to be chosen: the integration radius and the resolution range to be used for Patterson calculation. In the rotation function only intra-molecular vectors need to be considered. Since all vectors in a Patterson start at the unit cell axes origin, the vectors closest to the origin will in general be intra-molecular. By judiciously choosing a maximum Patterson radius, we can improve the chances of finding a strong rotation hit. Usually a value of the same order of magnitude of the search probe dimensions is chosen. Regarding the second parameter, high resolution reflections (above 3.5 Å) will differ a lot because there are related to the residue conformations. On the other hand, low resolution reflections, below 10Å are influenced by crystal packing and the solvent arrangement. Thus, the resolution range that should be used is usually within 10-3.5Å.

4. Output files. The output files to check are the file .log that lists all the operations performed by the program and the coordinates file representing the MR solution, *i.e.* the model rotated and translated in the real cell, in pdb format. As shown in **Table II**, after the rotational and translational searches performed, the program lists all the possible solutions (see **Note 16**) followed by the rotation angles and the translation shifts necessary to position the model in the real cell, the crystallographic R factors and finally the correlation coefficients (see **Note 17**). The first line of the **Table II** represents a clear solution for a MR problem. In fact the crystallographic R factor is below 0.5, and the correlation coefficient is very high (75.9 %). Moreover, there is a jump between the first possible solution (first line) and the second possible solution (second line).

3.6 Structure refinement

Several programs can be used to perform structure refinement. The most common are CNS written by Axel Brünger (**19**) that uses conventional least square refinement as well as simulated annealing to refine the structure and REFMAC5 (CCP4i suite) written by Murshudov (**20**) that uses the maximum likelihood refinement. Although CNS and many other programs have been used with success, in this chapter we illustrate the use of REFMAC5 implemented in CCP4i in that it provides a graphic interface to compile the input files which is particularly helpful for beginners.

1. Rigid body refinement. Firstly, the initial position of the molecules in the unit cell and in the crystal cell provided by MR procedures has to be refined. For this purpose Rigid Body refinement should be performed. This method assigns a rigid geometry to parts of the structure and the parameters of these constrained parts are refined rather than individual atomic parameters. The input files to be uploaded are the MR solution and the .mtz file containing the experimental reflections. The resolution to run rigid body refinement has to be chosen (in general the rigid body refinement should start at the lowest resolution range) and the rigid entity should be defined (this can be an entire protein a protein subunit or a protein domain). The definition of the rigid entity in REFMAC5 consist simply in selecting the chain and the protein region that is to be fixed.

2. Output files. The output files are: (a) the .log file that contains a list of all the operations performed, statistics about the geometrical parameters after each refinement cycle, crystallographic R factor and R free factor values, and finally the figure of merit (see **Note 18**); (b) the .pdb file containing the refined coordinates of the model; (c) the .mtz file containing the observed structure factors (F_{obs}), the structure factor amplitudes calculated from the model (F_{calc}) and the phase angles calculated from the model.

3. Coordinates and B factors refinement. The program REFMAC 5 refines the x, y, z and B parameters using the maximum likelihood method. As for the Rigid Body refinement the input files are the .mtz file containing the F_{obs} and the .pdb file containing the coordinates of the model. It is necessary to restrain the stereochemical parameters also using the maximum likelihood method. It is possible to choose a numerical value for the relative weighting terms or, more easily, to choose a single value for the so called "weight matrix" that allows the program to restrain all the stereochemical parameters together. The value of the "weight matrix" should be between 0.5, indicating loose stereochemical restraints, and 0, indicating strong stereochemical restraints which keep geometrical parameters of the macromolecules near the ideal values. In REFMAC5 NCS (see **Note 19**) restraints can be also used for refinement.

3.7 Model building

After the Molecular Replacement and the first cycles of coordinates refinement, only a partial model has been obtained. In this model the side chains are absent, and often parts of the model not match with the electronic density map. Therefore, the building of the first structural elements is followed by refinement

cycles that should lead to an improvement on the statistics *i.e* the R factor has to decrease and the figure of merit has to increase. The most common programs used for model building are QUANTA, O (**21**) COOT (**22**), and XFIT (XTALVIEW PACKAGE) (**23**). XFIT and COOT permit to directly calculate density maps. Two maps are necessary to build a model: the 2Fo-Fc map contoured at 1σ which is used to trace the model and the Fo-Fc map contoured at 3σ, which is necessary to observe the differences between the model and the experimental data.

1. Starting point. The first thing to do is find a match between protein sequence and 2Fo-Fc density map. If the phases are good, this operation is not too difficult. The electron density map is clear (especially if it has been calculated from high resolution data) and allows the identification of the aminoacids (see **Note 20**) (**Figure 4**).

2. Initial building. Once the first residue has been identified and fitted into the electron density map, model building can be performed by fitting the whole protein sequence residue by residue in the map (see **Note 21**).

3. Building of the unfitted structure elements. If the initial model does not contain all the protein residues, it is possible to build the main chain of the protein region missing from the model "ab initio", using one of the programs cited above. As an example, with XFIT it is possible to add to the model Cα atoms after or before a selected residue. After a Cα is inserted in the electron density map in a right position it is possible to substitute the Cα with the desired aminoacid, which is automatically bound to the rest of the protein. The main chain of the missing protein region can also be constructed using the program database after a suitable piece of structure has been built.

4. Omit map. If a part of the structure does not match the map, this means that it is built incorrectly. Thus, it is possible to use, as major strategy, for overcoming phase bias, the omit maps. In practice, the model region that has to be refitted is removed and the maps are recalculated, after a few refinement cycles. This method allows the phases calculated from the rest of the model to phase the area of interest with no bias from the model left out.

5. Optimization. At this stage, large sections of the structure should be approximately fitting the electron density map (**Figure 5**). The next step is the choice of the correct side chain rotamers. This operation may be done by hand or by using real space refinement tools. Finally water molecules and ions and/or ligands bound

to the protein have to be identified and added to the model. For this purpose only the Fo-Fc map contoured at 3σ is used. The water molecules can be added either manually or automatically.

4. Notes

1. The intercepts of the planes with the cell edges must be fractions of the cell edge. Therefore, cell intercepts can be at 1/0 (= ∞), 1/1, 1/2, 1/3 … 1/n. The conventional way of identifying these sets of planes is by using three integers that are the denominators of the intercepts along the 3 axes of the unit cell, hkl, called Miller indices. If a set of planes had intercepts at 1/2, 1/3, 1/1 then the planes would be referred to as the (2 3 1) set of planes.
2. Another advantage of synchrotron radiation is its tunability, which allows the user to select radiation wavelengths higher or lower than 1.5418 Å (copper radiation). Collection of data at wavelengths below 1.5418Å results in a lower signal to noise ratio.
3. A crystal can be regarded as a three-dimensional grid and you can imagine that this will produce a three-dimensional X-ray diffraction pattern. As with electron microscope grids, the pattern is reciprocal to the crystal lattice. The planes that intersect the sphere in **Figure 1**. are layers in a three-dimensional lattice, called reciprocal lattice because the distances are related reciprocally to the unit cell dimensions. Each reciprocal lattice point corresponds to one diffracted reflection. The reciprocal lattice is an imaginary but extremely convenient concept to determine the direction of the diffracted beams. If the crystal rotates, the reciprocal lattice rotates with it. In an X-ray diffraction experiment the direction of the diffracted beams depends on two factors: the unit cell distances in the crystal, from which the unit cell distances in the reciprocal lattice are derived, and the X-ray wavelength. As indicated in the **Figure 1**, diffraction conditions are determined not only by the reciprocal lattice but also by the radius of the sphere of the reflection or "Ewald sphere", of which radius is 1/λ.
4. The imaging plate detector is formed by a photosensitive plate, made of BaFBr:Eu. When hit by a radiation, the plate produces a latent image that can be excited by a laser, operating at 633 nm, which generates a 390 nm radiation corresponding to the fluorescence transition of Europium. This radiation is then collected in the photomultiplier and converted to an electric signal.

5. The CCD camera (Charged Coupled Device) is another kind of area detector. The detector surface is constituted by voltage sensitive elements (pixels). They have a high dynamic range, combined with excellent spatial resolution, low noise and high maximum count rate.
6. The resolution is defined as the minimum inter-planar spacing of the real lattice for the corresponding reciprocal lattice points (reflections) that are being measured. It is directly related to the optical definition in which it is the minimum distance that two objects can be apart and still be seen as two separate objects. Thus, high resolution means low minimum spacing. Resolution is normally quoted in Ångstroms (Å).
7. An oscillation image (also called frame) is obtained by rotating a crystal continuously through 0.2-1.0° about a fixed axis, called ϕ axis, perpendicular to the incident X-ray beam.
8. The Patterson function is a Fourier summation with intensity as coefficient and without phase angle.This can be written as: $P(u,v,w)=\Sigma|F(hkl)|^2\cos 2\pi(hu+kv+lw)$. Further, it can be demonstrated that the Patterson function can be alternatively written as the self-convolution of the electronic density: $P(u,v,w)=\int_r \rho(r)\rho(r+u)dr$.
9. Macromolecules in crystals are not static. Atoms vibrate around an equilibrium position and, as a consequence, the intensity of the diffracted beams are weakened. This phenomenon is expressed by the temperature factor $B=8\pi^2\times u^2$ where "u" is the mean square displacement of atoms around the atomic positions.
10. The maximum likelihood method, based on the Bayes theorem, assumes that our set of measurements is the most probable. On this basis, the most appropriate value for each variable (for example bond distances, angles, etc.) is that maximizing the probability P to observe all of our measurements.
11. Protein crystals are affected by lattice defects. Therefore, they are formed by different mosaic blocks with slightly different orientations. As an ideal single crystal has a mosaicity equal to 0 degrees, a good quality protein crystal should have a low mosaicity (between 0.2 and 0.5 degrees).
12. Table I shows the output of the program Denzo after the autoindexing. In this table, all the 14 possible Bravais lattices are listed from the highest symmetry (primitive cubic) to the lowest (primitive triclinic) and allows the intentification of the crystal lattice. After the lattice name the table display a percentage value that represents the amount of distortion that unit cell parameters would suffer in order to fit the lattice. Next to this percentage the "distorted-to-fit" unit cell parameters are listed. Below this values, the

undistorted unit cell parameters are shown for comparison. The goal of the autoindexing procedure is to find the highest symmetry lattice which fits the data with minimal distortion. In the shown example, the crystal lattice is primitive hexagonal since 0.22% is an accetable amount of distortion, expecially given that the unit cell parameters were refined from a single frame. The crystal lattice should be confirmed by the overall Denzo data reduction and Scalepack scaling procedure.

13. χ^2 is a parameter related to the ratio between intensity and its standard deviation σ for all measurements and its value should be around 1. The χ^2 is mathematically represented by the following equation:

$$\chi^2 = \frac{\sum_{ij}\left(\left|I_{ij}(hkl) - \left\langle I_i(hkl)\right\rangle\right|\right)^2}{\sigma i^2 \frac{N}{N-1}}$$

where hkl are the Miller indices and N indicates the number of observations.

14. R_{sym} is the parameter used to compare the intensity (I) of symmetry related reflections for *n* independent observations:

$$R_{sym} = \frac{\sum_{hkl}\sum_{i}\left|I_i(hkl) - \overline{I(hkl)}\right|}{\sum_{hkl}\sum_{i} I_i(hkl)}$$

the index *i* indicates the experimental observations of a given reflection. $\overline{I(hkl)}$ is the average intensity for symmetry related observations.

15. To avoid model bias often the model is transformed in a poli-Ala search probe. Only the coordinates of the polipeptide backbone and of Cβ atoms are conserved, whereas the side chains atoms are deleted .

16. The MolRep solutions represent the highest superposition peaks between the experimental Patterson and the Patterson calculated from the search probe, rotated and translated in the real cell.

17. Correlation coefficient. The correlation coefficient value (CC_f) is comprised between 0 and 1 and measures the agreement between the structure factors calculated from the rotated and translated model and the observed structure factors. The correlation coefficient is calculated by REFMAC5 using the following formula:

$$CCf = \frac{\left[\sum_{hkl}\left(|Fobs||Fcalc|\right)-\left(\langle|Fobs|\rangle\langle|Fcalc|\rangle\right)\right]}{\left[\sum_{hkl}\left(Fobs^2-\langle Fobs\rangle^2\right)\left(\sum_{hkl}\left(Fcalc^2-\langle Fcalc\rangle^2\right)\right)\right]^{1/2}}$$

18. Figure of merit. The "figure of merit " m is: $m = \frac{|F(hkl)best|}{|F(hkl)|}$

where: $\mathbf{F}(hkl)_{best} = \frac{\sum_{\alpha} P(\alpha)\mathbf{F}hkl(\alpha)}{\sum_{\alpha} P(\alpha)}$

P(α) is the probability distribution for the phase angle α and F_{hkl}(best) represents the best value for the structure factors. The m value is comprised between 0 and 1 and is a measure of the agreement between the structure factors calculated on the basis of the model and the observed structure factors. If the model is correct the figure of merit approaches the value of 1.

19. Non crystallographic symmetry (NCS) occurs when the asymmetric unit is formed by two or more identical subunits. The presence of this additional symmetry could help to improve the initial phases and obtain interpretable maps for model building using the so-called density modification techniques (**24**).

20. Usually the sequence regions that contains the largest number of aromatic residues is chosen to start the search. The aromatic residues (especially tryptophan) contains a high number of electrons and display an electronic density shape easy to recognise (see **Figure 4**).

21. All the mentioned programs (XFIT, O, QUANTA, etc.) are provided with functions that identify the maxima in the Fo-Fc map above a given treshold (usually 3σ is used) and places the water molecules at the maxima peaks.

22. DpsTe. DpsTe is a member of the Dps family protein (DNA binding protein from starved cells). DpsTe has been isolated and purified from the cyano bacterium *Thermosynechococcus elongatus*. The structure has been solved by Molecular Replacement at 1.81 Å resolution and has been deposited in the Protein Data bank with the accession number 2C41.

References

1. Ducruix, A., and Giegè, R., (1992), *Crystallization of Nucleic Acids and Proteins*, (Rickwood D. and Hames B. D., ed.), Oxford University Press, New York, NY, pp. 7--10.

2. Hahn, T. (ed.) (2002) *International table of crystallography*, Kluwer Academic Publishers, Dodrecht.

3. McPherson, A. J. (1990) Current approach to macromolecular crystallization. *Eur. J. Biochem*. **189**, 1--23.

4. Matthews, B.W (1968) Solvent content of protein crystals. *J. Mol. Biol*., **33**, 491--7.

5. Friedrich, W., Knipping P., and Laue, M., (1981), *Structural Crystallography in Chemistry and Biology*, (J. P. Glusker ed.), Hutchinson & Ross:Stroudsburg, PA, pp 23--39.

6. Bragg, W. L. and Bragg, W. H. (1913) The structure of crystals as indicated by their diffraction of X-ray . *Proc. Roy. Soc. London* **89**, 248--77.

7.Chothia C., Lesk A. M. (1986) The relation between the divergence of sequence and structure in proteins. *EMBO J*. **5**, 823--6.

8. Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., Bourne P.E., (2000) The Protein Data Bank. *Nucl Ac. Res*., **28**, 235--242.

9. Crowther, R. A. (1972) *The Molecular Replacement Method*, (Rossmann, M.G., ed.) Gordon & Breach, New York, NY, pp. 173--178.

10. Hendrickson, W. A. (1985) Stereochemically restrained refinement of macromolecular structures. *Methods Enzymol.* **115**, 252--270.

11. Brunger, A. T., Adam,s P. D. and Rice, L. M. (1999) Annealing in crystallography: a powerful optimization tool. *Prog Biophys Mol Biol*. **72**, 135--55.

12. Rodgers, D. W., Rodgers D. W. (1994) Cryocrystallography. *Structure* **2,** 1135--40.

13. Otwinoski, Z. and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276,** 307--326.

14. CCP4 (Collaborative Computational Project , number 4) (1994) The CCP4 suite: programs for protein crystallography. *Acta Cryst.* **D50**, 760--763.

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

16. Altshul S. F. and Koonin, E. V. (1998) Iterated profile searches with PSI-BLAST - a tool for discovery in protein databases. *TIBS* **23**, 444--7.

17. Navaza, G. (1994) AMORE: an automated package for molecular replacement. *Acta Crystallogr.* **A50**, 157--163.

18. Vagin, A., and Teplyakov, A. (1997) MOLREP: an automated program for molecular replacement. *J. Appl. Crystallogr.* **30**, 1022--1025.

19. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. and Warren, G. L. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr* **D54**, 905--21.

20. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method . *Acta Cryst.* **D53**, 240--255.

21. Emsley, P. and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Cryst*. **D60**, 2126--2132.

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

23. McRee, D.E. (ed.) (1993) Practical protein crystallography. Academic Press, Inc., San Diego, Ca.

24. Rossmann, M.G. and Blow, D.M. (1962) The detection of subunits within the crystallographic asymmetric unit. *Acta Cryst*. **15**, 24--31.

25. Ilari A., Johnson K. A., Nastopoulos V., Verzili D., Zamparelli, C.,Colotti, G., Tsernoglou D. and Chiancone E. (2002) The crystal structure of the sorcin calcium binding domain provides a model of Ca^{2+}-dependent processes in the full-length protein. *J. Mol Biol* **317**, 447--458.

Figure legends

Figure 1. Ewald sphere. A diagramatic representation of the generation of an X-ray diffraction pattern (see **Note 3**).

Figure 2. a. "Hanging drop" crystallization method. A drop of protein solution is suspended from a glass cover slip above a reservoir solution, containing the precipitant agent. The glass slip is siliconized to prevent spreading of the drop **b.** "Sitting drop" crystallization method. A drop of protein solution is placed in a plastic support, above the reservoir solution.

Figure 3. Diffraction oscillation image visualized with the program Xdisp (HKL2000 suite) of the whole human sorcin collected at the ESRF synchrotron radiation source (Grenoble, FR). The spot distances from the image centre are proportional to the resolution, so the spots at the image edge are the highest resolution spots.

Figure 4. Initial Electronic density map of Dps from *Thermosynechococcus Elongatus* (see **Note 21**) calculated after Molecular Replacement. Cα trace of the model is superimposed on the map. The electronic density of a Trp residue and a Tyr residue are easily recognizable in the map.

Figure 5. Electronic density map contoured at 1.0 σ of Dps from *Thermosynechococcus Elongatus* (see **Note 21**) calculated after many REFMAC5 refinement cycles. The final structure (thick lines) solved at 1.8 Å resolution is superimposed on the map.

Lattice	Metric tensor distortion index	Best cell (symmetrized) Best cell (without symmetry restrains)					
primitive cubic	60.04%	64.84	64.96	315.85	90.09	89.92	60.43
		148.55	148.55	148.55	90.00	90.00	90.00
I centred cubic	74.97%	65.32	322.35	322.36	20.04	84.33	84.38
		236.68	236.68	236.68	90.00	90.00	90.00
F centred cubic	79.51%	335.11	322.36	322.56	23.24	157.19	157.36
		326.68	326.68	326.68	90.00	90.00	90.00
primitive rhombohedral	2.75%	322.35	322.34	315.85	11.69	11.60	11.57
		320.18	320.18	320.18	11.62	11.62	11.62
		64.90	64.90	953.96	90.00	90.00	120.00
primitive hexagonal	**0.22%**	65.32	64.84	315.85	89.92	90.17	120.12
		65.08	65.08	315.85	90.00	90.00	120.00
primitive tetragonal	13.37%	64.84	64.96	315.85	90.09	89.92	60.43
		64.90	64.90	315.85	90.00	90.00	90.00
I centred tetragonal	13.68%	64.84	64.96	634.88	87.11	92.88	60.43
		64.90	64.90	634.88	90.00	90.00	90.00
primitive orthorhombic	13.37%	64.84	64.96	315.85	90.09	89.92	60.43
		64.84	64.96	315.85	90.00	90.00	90.00
C centred orthorhombic	0.09%	65.32	112.17	315.85	90.01	89.83	90.13
		65.32	112.17	315.85	90.00	90.00	90.00
I centred orthorhombic	13.68%	64.84	64.96	634.88	87.11	92.88	60.43
		64.84	64.96	634.88	90.00	90.00	90.00
F centred orthorhombic	2.37%	65.32	112.17	634.88	89.99	95.73	90.13
		65.32	112.17	634.88	90.00	90.00	90.00
primitive monoclinic	0.07%	64.84	315.85	64.96	90.09	119.57	90.08
		64.84	315.85	64.96	90.00	119.57	90.00
C centred monoclinic	0.05%	65.32	112.17	315.85	89.99	90.17	90.13
		65.32	112.17	315.85	90.00	90.17	90.00
primitive triclinic	0.00%	64.84	64.96	315.85	90.09	90.08	119.57
autoindex unit cell		65.24	65.24	315.85	90.00	90.00	120.00
crystal rotx, roty, rotz		-8.400	55.089	70.885			
Autoindex Xbeam, Ybeam		94.28	94.90				

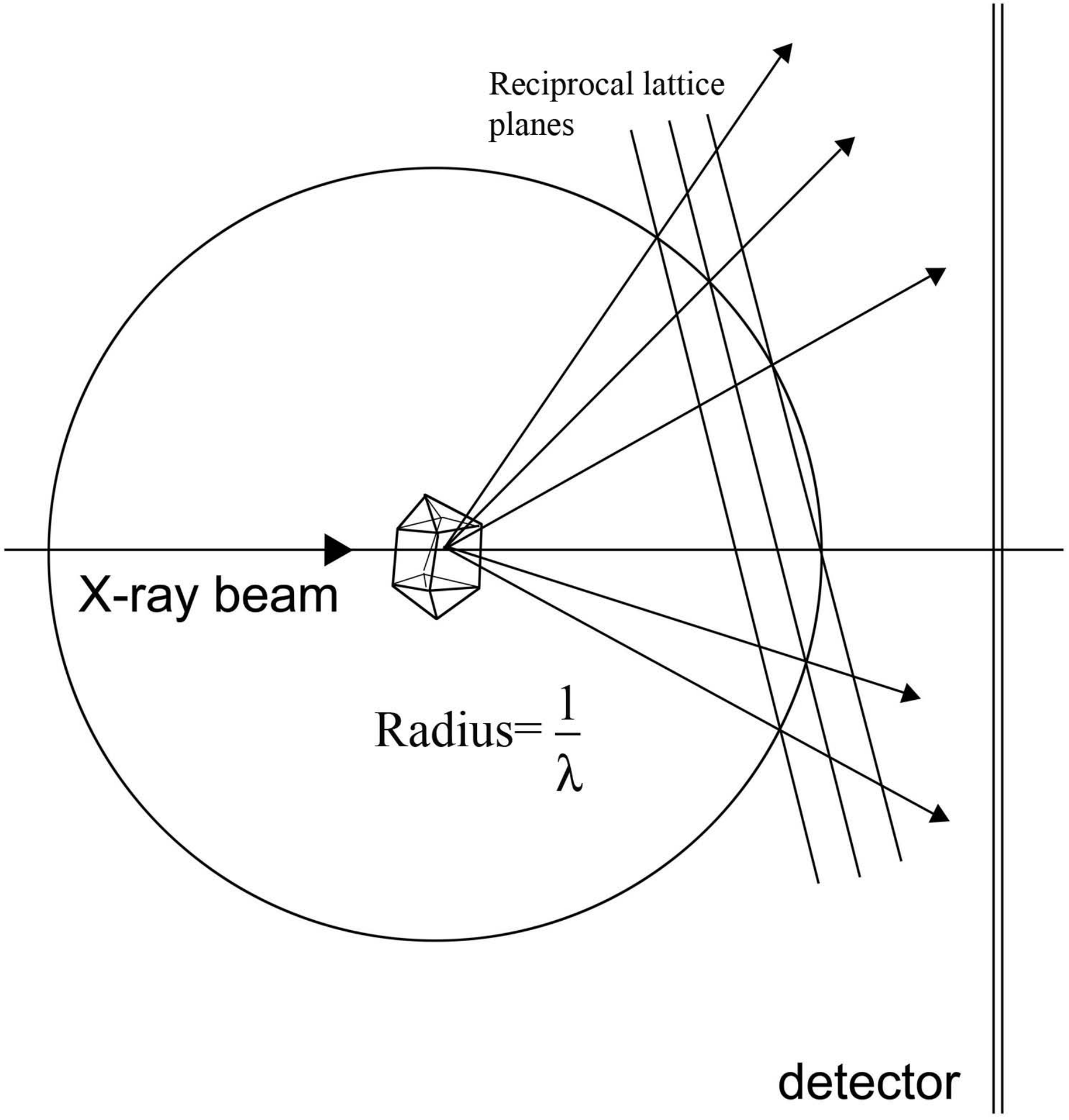
Table I. Output of the Denzo autoindexing routine. The lattice and unit cell distortion table, and the crystal orientation parameters are shown.The present results are obtained for the human sorcin (soluble Resistance related calcium binding protein) (**25**).

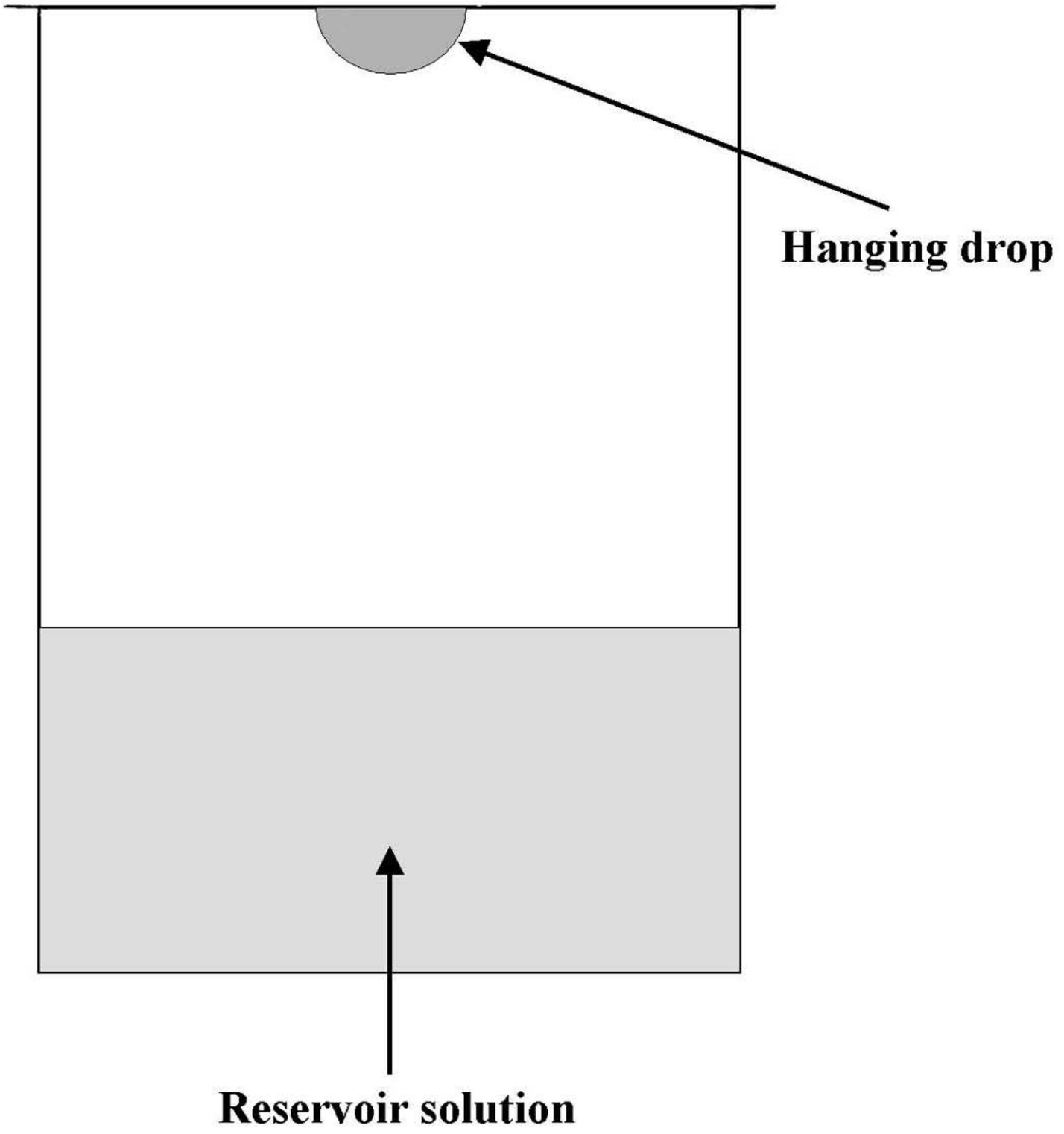
		alpha	beta	gamma	Xfrac	Yfrac	Zfrac	TF/sig	R-fac	Corr
Sol_TF_7	**1**	**32.27**	**84.87**	**78.84**	**0.824**	**0.498**	**0.091**	**65.34**	**0.333**	**0.759**
Sol_TF_7	2	32.27	84.87	78.84	0.324	0.041	0.092	25.76	0.482	0.478
Sol_TF_7	3	32.27	84.87	78.84	0.324	0.454	0.091	24.18	0.477	0.481
Sol_TF_7	4	32.27	84.87	78.84	0.324	0.498	0.016	23.57	0.483	0.467
Sol_TF_7	5	32.27	84.87	78.84	0.422	0.498	0.091	23.37	0.479	0.478
Sol_TF_7	6	32.27	84.87	78.84	0.324	0.498	0.325	23.12	0.482	0.471
Sol_TF_7	7	32.27	84.87	78.84	0.238	0.498	0.092	23.01	0.481	0.473
Sol_TF_7	8	32.27	84.87	78.84	0.324	0.498	0.372	22.99	0.479	0.475
Sol_TF_7	9	32.27	84.87	78.84	0.324	0.498	0.400	22.97	0.480	0.473
Sol_TF_7	10	32.27	84.87	78.84	0.324	0.000	0.196	22.93	0.490	0.456

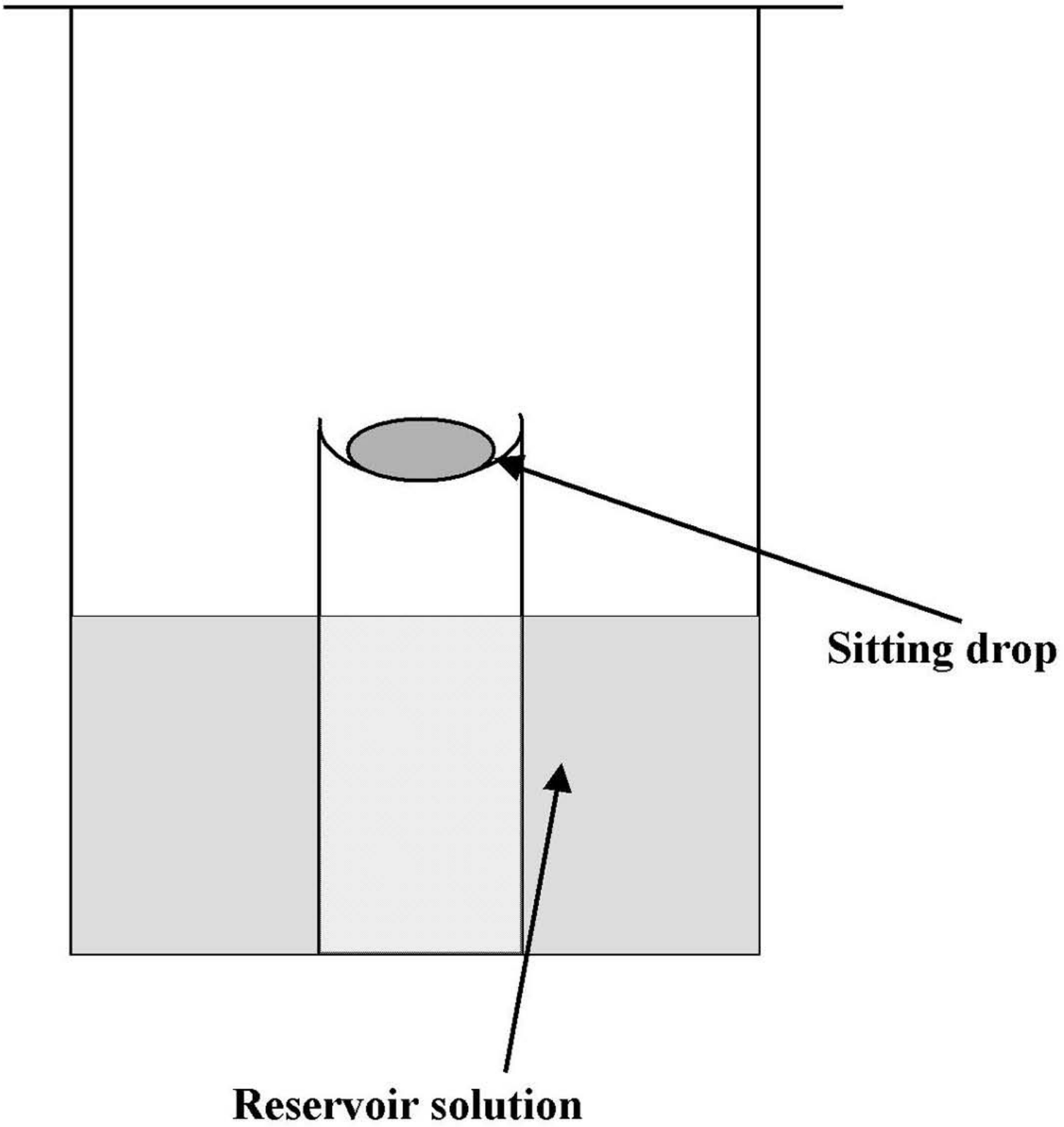
Table II. Output of the MolRep program after rotation and translation searches. The present results (data not published) have been obtained for the protein Dps (Dna binding proteins, see **Note 21**) from *Listeria monocitogenes* using as search model the Dps from *Listeria innocua* (Pdb code 1QHG)

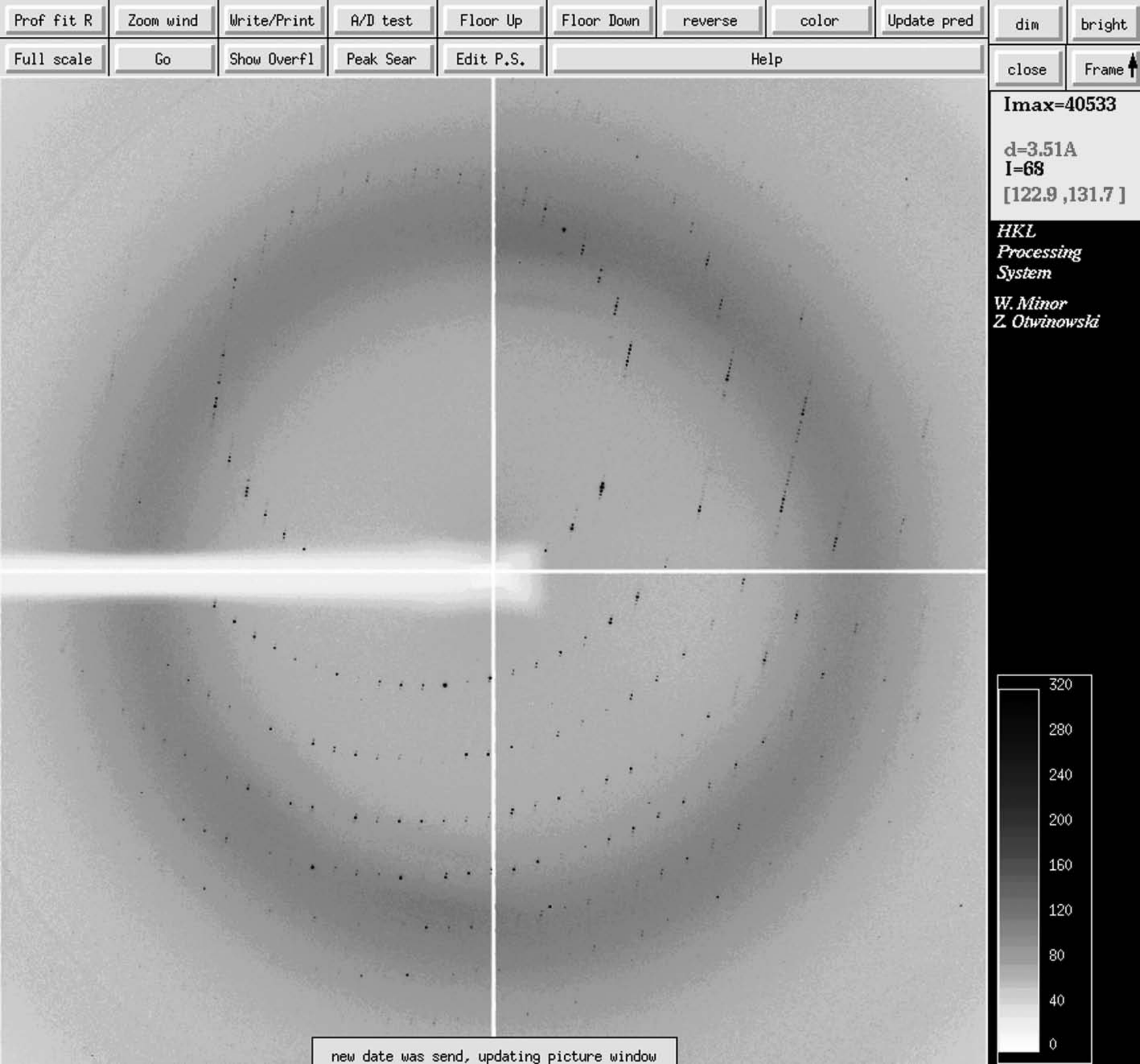
Ncyc	Rfact	Rfree	FOM	LLG	rmsBOND	rmsANGLE	rmsCHIRAL
0	0.213	0.213	0.862	1165259.2	0.004	0.734	0.055
1	0.196	0.210	0.865	1151022.5	0.010	1.022	0.074
2	0.191	0.209	0.867	1146576.9	0.011	1.106	0.080
3	0.188	0.209	0.868	1144297.8	0.011	1.144	0.083
4	0.187	0.209	0.869	1142920.2	0.011	1.166	0.085
5	0.186	0.209	0.870	1142088.8	0.011	1.178	0.086
6	0.185	0.209	0.870	1141496.4	0.011	1.186	0.087
7	0.184	0.209	0.870	1141031.5	0.011	1.190	0.088
8	0.184	0.209	0.871	1140743.6	0.011	1.192	0.088
9	0.184	0.209	0.871	1140461.8	0.011	1.195	0.088
10	0.183	0.209	0.871	1140311.0	0.011	1.196	0.088

Table III . Summary of 10 cycles of DpsTe (see **Note 21**) coordinate refinement using REFMAC5. The R_{fact}, R_{free}, Figures of Merits (FOM) and root mean square deviation values of some stereo-chemical parameters are shown.


Reciprocal lattice
planes
X-ray beam
Radius= $\frac{1}{\lambda}$
detector


Hanging drop
Reservoir solution


Sitting drop
Reservoir solution


Prof fit R
Zoom wind
Write/Print
A/D test
Floor Up
Floor Down
reverse
color
Update pred
dim
bright
Full scale
Go
Show Overfl
Peak Sear
Edit P.S.
Help
close
Frame
Imax=40533
d=3.51A
I=68
[122.9 ,131.7]
HKL
Processing
System
W. Minor
Z. Otwinowski
320
280
240
200
160
120
80
40
0
new date was send, updating picture window