

Practical applications of synchrotron radiation in the determination of bio-macromolecule three-dimensional structures

M. Nardini and M. Bolognesi

Department of Biomolecular Sciences and Biotechnology,
University of Milano, via Celoria 26, 20133 Milano, Italy

The last decade has witnessed the development of new methodologies in molecular biology and biochemistry which, together with the widespread availability of computing power for fast calculation and graphics, have supported the striking growth of crystallographic studies of biological relevance. This surge in new methodologies experienced by macromolecular crystallography has been largely fuelled by developments in structural genomics and by the needs of pharmaceutical industry. A key contribution to such rapid growth was certainly provided by the increased access to synchrotron sources, which resulted in new ways of collecting X-ray diffraction data and allowed novel structure determination techniques, otherwise not possible on conventional X-ray sources. Open access to synchrotron sources, together with the use of efficient X-ray detectors, is probably the fundamental reason of the exponential growth of structural biology studies during the last fifteen years.

The next paragraphs cover some general aspects and select recent advances in macromolecular crystallography, focussing especially on synchrotron radiation applications.

1. Macromolecular crystals

Almost every week new three-dimensional structures of proteins, nucleic acids, and their complexes appear in the most important scientific journals. The complexity of such structures is steadily growing, and their experimental resolution is often at atomic level. These results provide one of the most exciting and accurate tools for innovation in life sciences, particularly for applications in drug-design, immunology, virology, and in enzymology [1-5]. Protein crystallography, which has a growing role in the human genome project, is one of the most powerful techniques in modern biology for three-dimensional structure determination. However, it can only be applied providing suitable crystals can be obtained. The ability to produce suitable crystals is currently the major bottleneck to structure determination.

1.1 *Physical and chemical properties*

A crystal of organic material (such as proteins, DNA fragments, tRNA molecules, viruses, etc) is a three-dimensional periodic array of macromolecules. In terms of crystal size bio-macromolecular crystals are rather small, with volumes rarely exceeding 0.1 mm^3 and crystal lattice periodicity often $>100 \text{ \AA}$. Macromolecular crystals show unique physical and chemical properties, among other crystals in Nature. In particular, they display poor mechanical stability and a high solvent content. Both the fragility of these crystals and their sensitivity to external conditions are related to the weak interactions between macromolecules in the crystal lattice, which are based mainly on hydrogen bonds and van der Waals interactions. The energy responsible for crystal lattice stabilization is very low ($<10 \text{ kcal/mol}$), typically lower than the energy required for protein folding. Therefore, the lattice forces usually do not disrupt the overall conformation of the crystallized macromolecules. This is a critical point since it ensures that the three-dimensional structure of a macromolecule in a crystalline environment is identical to the structure of the same molecule in solution.

The fragility of macromolecular crystals is also related to their high solvent content (from 30% to more than 80% v/v). In fact, to prevent crystal cracking or collapse due to dehydration, the crystals must be kept in a solvent saturated environment. The high solvent content of the crystals has two important and useful implications. Firstly, the macromolecule concentration in the crystal is of the same order of magnitude of that typical of cytoplasmic compartments *in vivo* (about 10^{-2} M). Secondly, extended solvent channels (completely crossing the crystal volume and touching each crystallized protein molecule) allow the efficient diffusion of small molecules within the crystals. The diffused molecules can be either heavy atoms or other ligands (substrates, inhibitors, cofactors, etc), thus providing key information for structure determination purposes and for the mechanistic description of enzymatic reactions.

1.2 Crystal growth

Despite dramatic progress in macromolecular crystallography research, crystallization remains a major bottleneck in the structural characterization of macromolecules. In fact, the methods used to produce crystals have also evolved over the years as a result of increased understanding and advancing technology, but crystallization strategies continue to be rooted in massive trial-and-error approaches [6-8]. In general, it is common belief that insufficient purity of the sample is the most probable cause of unsuccessful crystallization. The sample quality requirements for protein crystallography are more stringent than the requirements for most biochemical tests. To grow macromolecular crystals of good quality the protein sample must be highly homogeneous: other compounds should be absent, all protein molecules should have the same surface properties (particularly for what concerns surface charge distribution), and the same oligomerization state.

The fundamental task in protein crystallization is to create a solution supersaturated with protein that will produce single, well-ordered crystals. More often than not, supersaturated solutions produce precipitate or phase separation instead. There is no *a priori* theory for discerning which solutions will produce crystals and which will produce precipitate. So the crystallization process is separated into two stages: screening and optimizing. Each stage is conducted with a largely empirical approach. The screening process discovers lead crystallization conditions or 'hits', which typically produce microcrystals, thin rods or plates. The optimization process is devoted to grow crystals of size and regularity adequate for a successful X-ray diffraction data collection.

Crystal growth is a multiparametric process involving three basic steps: nucleation, growth, and cessation of growth (Fig. 1). To achieve the spontaneous formation of the first ordered aggregates (nuclei) the sample solution is brought to supersaturation. Supersaturation can be achieved by varying intrinsic physical and chemical parameters, such as ionic strength, dielectric constant of the solvent, pH, temperature, macromolecule concentration, concentration of chemicals (precipitant, buffer, additives). Once nuclei have formed, crystal growth can begin, and proceed while the degree of supersaturation is reduced. Maintaining high supersaturation would result in the formation of too many crystallization nuclei, which will develop into too many small crystals. Furthermore, crystals should grow slowly to reach the maximum degree of intrinsic order. Cessation of growth can be triggered by different causes, such as growth defects, poisoning of the crystal faces, "ageing" (chemical reaction) of the macromolecules, or simply depletion of the macromolecules from the crystallization media.

From a practical point of view, precipitation of the macromolecules to form crystals is mostly achieved using techniques known as "vapour diffusion" (increase of protein concentration by evaporation of water in a sealed environment), or by dialysis (Fig. 1). A great number of experiments (100-500) are usually necessary to determine the best crystallization conditions. Therefore, crystallization experiments are run on a very small scale (μ l samples). Macromolecular crystals suitable

for X-ray diffraction experiments have a typical linear size of 10-100 μm , and grow in a time period varying from few hours, to weeks or even months. In case of limited crystal growth, not related to biochemical problems, crystallization experiments carried out in space, under “microgravity” conditions, might improve the quality of the crystals [9].

When screening methods for protein crystallization fail, the modification of target proteins is often necessary to increase the susceptibility of the proteins to crystallization. Each change to the sequence of the protein (amino acid mutation) or its ligation state generates a new crystallization target for screening. This approach includes the crystallization of single domains in multi-domain proteins, of truncated forms at the N- and C-terminus, of truncated forms generated by limited proteolysis, and of single/multiple site-specific mutants. Site-directed mutagenesis is a common method to alter the surface properties of the target, or to stabilize the target. In addition, one can also attempt to engineer crystal contacts when the crystal structure of a homologous protein or a low-resolution structure of the target is available.

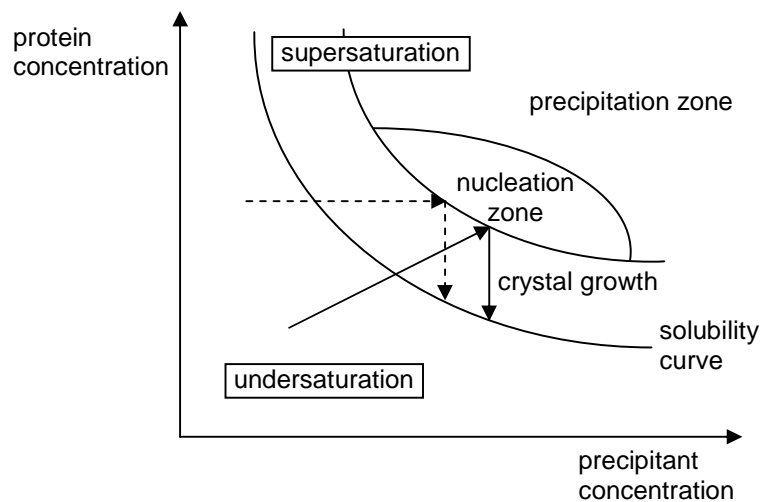


Figure 1. Solubility curve for a protein *versus* precipitant concentration (salt or another chemical parameter affecting protein molecule aggregation) in a typical vapour diffusion (continuous line) or dialysis (dashed line) crystallization experiment.

1.3 Crystal handling *ad* radiation damage

Macromolecular crystals are usually very fragile, and need special handling care. This unique property reflects the fact that the macromolecules are loosely packed in the crystal, with large solvent-filled channels crossing the crystal volume. Removal of the solvent destabilizes the crystal, which can easily dissolve or shatter. Protein crystals must, therefore, always be kept in their mother liquor or in its saturated vapour, even when exposed to X-rays. For data collection purpose, the crystal can be hosted in a thin-walled capillary, containing a small amount of mother liquor. Water vapour will guarantee proper hydration of the crystal when the capillary is sealed on both sides by resin (Fig. 2A). Then, the capillary is easily mounted on a goniometer head and properly oriented for X-ray exposure. Unfortunately, the energy released by absorption of X-rays in a crystal inevitably damages it by causing the formation of free radicals in the crystal (especially $\text{OH}\cdot$ due to water radiolysis) which react

with the protein molecules and irreversibly damage their crystal packing. The most common symptoms of radiation damage are a decrease of diffraction intensity and resolution limits (ultimately resulting in the X-ray pattern dying away), a variation of the unit cell volume with consequent non-isomorphism within a data series (thus hampering traditional phasing methods), and site-specific damage. The latter occurs in a well-defined order, starting with the breakage of disulphide bonds, followed by decarboxylation of aspartates, glutamates and the C-terminus, and then loss of the hydroxyl group from tyrosines [10]. The extent of radiation damage depends on the applied radiation dose, on the photon energy (it is lower for energy higher than 12.4 keV, or $\lambda=1.0 \text{ \AA}$), on the type of proteins and on the chemical composition of the solvent. The damage continues even after X-ray exposure, due to the long half-life time of the free radicals.

Nowadays, the extensive use of synchrotron radiation has made almost obligatory the cooling of the crystal samples to cryogenic temperature. In liquid nitrogen at 100 K, the radiation damage effects are so much reduced that they are often negligible. Most protein crystals can be cooled to cryogenic temperatures (with appropriate care for the solvent properties) using the so called “flash-freezing” technique. The crystal is “fished” by using a nylon fibre loop, where it remains suspended in a thin film of solvent (Fig. 2B). The loop is then cooled very rapidly by immersion into liquid nitrogen through the temperature where ice crystals may grow, to below 190 K where water forms a stable glassy state. If flash-freezing is successful, the liquid film in the loop freezes into a glass and remains clear. For data collection, the loop is mounted onto a goniometer head, where it is held in a stream of cold nitrogen gas (Fig. 2C). Crystallographers can take or ship loop-mounted flash-frozen crystals to synchrotron, minimizing handling of crystals at the data collection site.

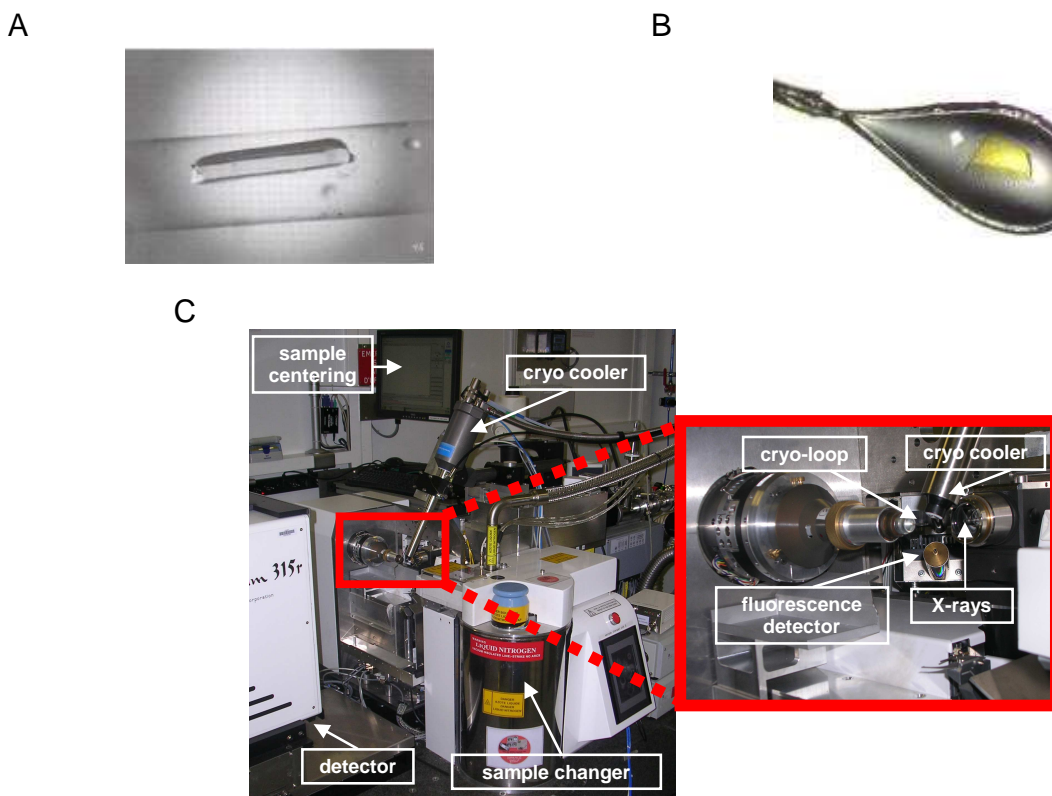


Figure 2. Crystal mounting. (A) Protein crystal in equilibrium with liquid solvent in a thin-walled capillary. (B) Crystal on a fibre loop at 100 K. (C) Typical set-up for recording data from a slowly rotated crystal at a MAD synchrotron beamline (ESRF, Grenoble, France).

2. X-ray diffraction and phasing problem

2.1 Fundamentals of crystallography

Let's now consider a crystal ideally schematised by a mosaic of crystalline blocks. At fixed wavelength, a Bragg reflection only occurs if the incident beam makes the correct angle with the reflecting planes. Because of the slight disorder in a real crystal, the reflection can be fully excited if the crystal is rotated through its reflecting position. In fact, all fixed-wavelength observing techniques use a rotating crystal system. Usually, the crystal is rotated about an axis orthogonal to the X-ray beam (Fig. 2C). For a protein crystal, a very large number of diffracted beams (called 'reflections' to remind Bragg's law) may be recorded, and it is necessary to avoid them overlapping. By rotating the crystal through a small angle (0.5-1°), they can easily be observed separately (Fig. 3). For macromolecules in a crystal to be observed at atomic resolution, very many reflections must be recorded. At the modest resolution of 2 Å a moderate-sized protein unit cell edge of 50 Å will give 25 orders of diffraction along this direction. Therefore, in three dimensions the hkl all vary between -25 and +25, assuming that the three cell dimensions are similar. Very roughly 125000 different Bragg reflections exist to this resolution and should be collected. In fact, the number of independent reflection is reduced by a factor of two if the Friedel's law applies and the number is further reduced depending on crystal symmetry.

If the crystal is rotated with an angular velocity ω through the reflection position, then the total energy of a diffracted X-ray along the direction defined by the Bragg reflection associated to the lattice planes hkl is expressed by the Darwin equation:

$$E_{hkl} = \frac{e^4}{m^2 c^4 \omega} I_0 \lambda^3 A L P \frac{V_x}{V_0^2} |\mathbf{F}_{hkl}|^2$$

where I_0 is the incident beam intensity, λ is the wavelength, A is the correction due to the absorption of the sample, L and P are the Lorentz and the polarization factors, V_x is the volume of the crystal and V_0 the volume of the unit cell. Each structure factor \mathbf{F}_{hkl} is a complete description of a diffracted X-ray recorded as reflection hkl and, therefore, is characterized by an amplitude and a phase and it can be represented in the imaginary plane as a vector (Fig. 4A).

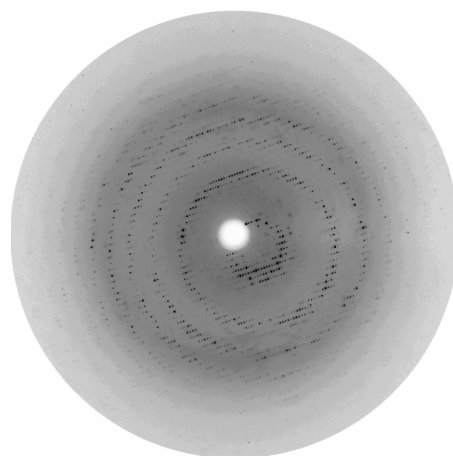


Figure 3. Diffraction pattern of a protein crystal subjected to small-angle oscillation data collection.

The Darwin equation shows that many different factors contribute to weaken the intensities of the waves diffracted by macromolecular crystals: the amplitude of the structure factors $|\mathbf{F}_{hkl}|$ are small, due to the low atomic number of the component elements (H, C, N, O, etc) of the biological macromolecule, the crystal volume is always small, while the unit cell is normally quite large. Therefore, significant diffracted intensities can be measured if high intensity incident radiation is applied, and if the sample X-ray absorption (by varying λ) as well as the diffused background radiation (by using a highly collimated beam) is minimized. All these experimental adjustments can be achieved using synchrotron radiation, while they are more difficult (or impossible) to achieve on a normal rotating anode X-ray source [11, 12].

A unique application of the synchrotron radiation in bio-crystallography is the so called Laue technique [13]. In this method the sample is hit by “white” radiation ($0.5 \text{ \AA} < \lambda < 3.0 \text{ \AA}$). As a consequence, many families of lattice planes are simultaneously “excited”, thus producing a complex pattern of diffraction that can, however, be deconvoluted. The advantage of this technique is that very short X-ray pulses ($\ll 1 \text{ s}$) can be used, due to the high intensity of the white radiation. In some cases (high symmetry of the crystal) one single pulse of radiation is sufficient to record the full diffraction diagram. This technique is at the basis of time-resolved crystallography experiments, which have found their major applications in the study of biological photoactivated protein reactions. The development of time-resolved multiwavelength Laue crystallography allows following the time course of biochemical processes (*i.e.* the catalytic cycle of an enzyme) through analysis of the related high resolution three-dimensional structure(s), with a time resolution in the (sub)nanosecond-millisecond range. In the past few years, this technique has evolved considerably due to the improvement of synchrotron sources, of beamline optics, charge-coupled detectors and of the optimized reaction triggering strategies based on femtosecond-picosecond laser pulses. In fact, the dynamical processes investigated with Laue diffraction require an accurate triggering event (initiated usually by a laser pulse) to achieve the correct timing between protein structural transitions and X-ray beam exposure [14-16].

2.2. Solution of the phase problem

In the diffraction experiment we measure the intensity of the diffracted beam I_{hkl} , which is proportional to the square of the amplitude of the structure factor \mathbf{F}_{hkl} . The structure factor is a function of the scattering due to the individual atoms comprising the unit cell electron density, f_j , known as the atomic scattering factor:

$$\mathbf{F}_{hkl} = \sum_{j=1}^N f_j \exp [2\pi i(hx_j + ky_j + lz_j)]$$

Instead of summing over all N separate atoms in the unit cell, we can integrate over all electrons in the unit cell volume V :

$$\mathbf{F}_{hkl} = \int_V \rho(xyz) \exp [2\pi i(hx + ky + lz)] dV$$

The reverse of this Fourier transform provides an expression of the electron density as a function of all structure factors:

$$\rho(x,y,z) = 1/V \sum_{hkl} \mathbf{F}_{hkl} \exp [-2\pi i(hx + ky + lz)] = 1/V \sum_{hkl} |\mathbf{F}_{hkl}| \exp (i\alpha_{hkl}) \exp [-2\pi i(hx + ky + lz)]$$

So, to calculate the electron density at the position (x,y,z) in the unit cell of a crystal requires us to perform the summation over all the *hkl* planes of the contributions to the point (x,y,z) of waves scattered from plane *hkl* whose amplitudes must be added with the correct relative phase relationship (α_{hkl}). We can measure the structure factor amplitudes, but the phases are lost in the experiment. This is “the phase problem” [17].

A similar equation can be calculated using the diffraction intensities as coefficients and with all phase angles equal to zero:

$$\rho^P(u,v,w) = 1/V \sum_{hkl} |F_{hkl}|^2 \cos[2\pi(hu+kv+lw)]$$

This equation is the Patterson function. It can be always calculated from the diffraction data since it depends on the squared structure factor amplitudes and it can be demonstrated to be a self-convolution of the electron density. The Patterson function is particularly useful both for the localization of “heavy atoms” (meaning atoms with atomic number higher than typical protein atoms), when heavy atom methods are applied for phasing, and for the correct orientation and translation positioning of a model in the crystal unit cell when the molecular replacement phasing method can be applied [13].

2.3 The MIR method

The Multiple Isomorphous Replacement (MIR) method is the most successful and general experimental approach for the determination of phase angles and their assignment to the corresponding structure factors. The method is based on the possibility of modifying structure factors and phases of the “native” protein crystal by specific binding of heavy atoms. When the position of the heavy atoms bound to the protein have been found by a Patterson map calculation, then it is possible to calculate the phase of each observed structure factor by proper consideration of the variation of the structure factor moduli in the native protein and in the derivatized crystals [13].

Practical application of the MIR method requires the comparison of X-ray diffraction patterns of the native protein crystal and of the crystals of at least two independent heavy atom derivatives. In case of perfect isomorphism (the conformation of the protein molecules, their position and orientation relative to the crystallographic axes, as well as the unit cell parameters must be exactly the same in the native and derivative crystals) the structure factor differences between the native protein crystals (\mathbf{F}_P) and the derivatives (\mathbf{F}_{PH}) are exclusively due to the contribution of the bound heavy atoms (\mathbf{F}_H). The contribution of the heavy atom replacement to the structure factor amplitude and phases is best illustrated on the Argand diagram (Fig. 4A).

The structure factor of the heavy atom derivative is then expressed by the vectorial summation:

$$\mathbf{F}_{PH} = \mathbf{F}_P + \mathbf{F}_H$$

The amplitudes $|F_{PH}|$ and $|F_P|$ are experimentally measured from the diffraction pattern of the native and derivative crystals, respectively. F_H (module and phase) is calculated from the position of the heavy atoms, which can be obtained from the Patterson synthesis of the isomorphous differences ($(|F_{PH}| - |F_P|)^2$). The native phase angle α_P (assigned to the structure factor \mathbf{F}_P) can be calculated by using the cosine rule:

$$\alpha_P = \alpha_H + \cos^{-1} [(|F_{PH}|^2 - |F_P|^2 - |F_H|^2) / 2|F_P||F_H|]$$

leading to two possible solutions symmetrically distributed about the heavy atom phase ($\alpha_P = \alpha_H \pm \alpha_0$). This two-fold ambiguity is better illustrated in the Harker construction (Fig. 4B). In this representation the possible structure factors \mathbf{F}_P and \mathbf{F}_{PH} are drawn by circles in the complex plane, having radius $|\mathbf{F}_P|$ and $|\mathbf{F}_{PH}|$, respectively. The two possible phase values occur at the circle intersections. The phase ambiguity can be removed by measuring at least two different heavy atom derivatives. The comparison of the two heavy atom contributions identifies the common correct value of α_P (Fig. 4C).

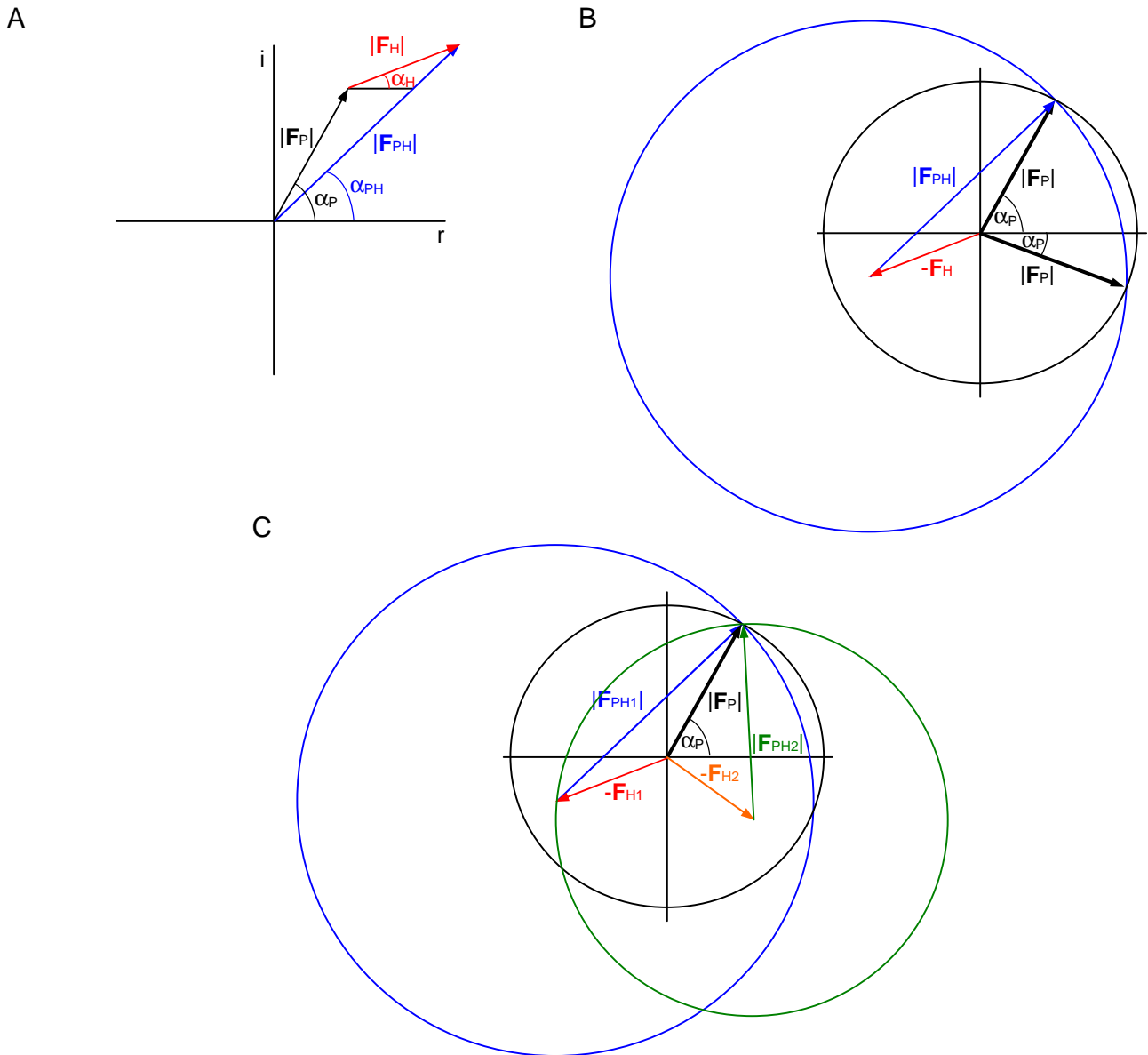


Figure 4. (A) Argand diagram for a single isomorphous replacement. Amplitude moduli and phases are indicated on the corresponding vectors for a native reflection (\mathbf{F}_P), a derivative reflection (\mathbf{F}_{PH}), and for the contribution of the bound heavy atom (\mathbf{F}_H). (B) Harker diagram for one heavy atom derivative. (C) Harker diagram for two heavy atom derivatives in a MIR experiment.

In real experimental cases, one heavy atom makes little difference to the structure factor amplitudes of the native protein. Furthermore, the calculation of \mathbf{F}_H is often not precise, due to an approximate deconvolution of the isomorphous difference Patterson, and due to the fact that binding of heavy atoms can introduce non-isomorphism between native and derivative crystals. As a result of several sources of experimental noise, the Harker circles do not intersect at a single point and uncertainty is present in the assignment of the correct phase angle α_P . A method recognising these difficulties has been therefore proposed, based on the consideration of a probability function for the phase angle α_P as a function of the observed experimental parameters. These phase probability distributions can be used to estimate “best phases”, with appropriate weighting, yielding the protein electron density map with the least error. If ε is the “lack of closure” of the phase triangle $\mathbf{F}_{PH} = \mathbf{F}_P + \mathbf{F}_H$ for the phase angle α

$$\varepsilon = |\mathbf{F}_{PH}|^{\text{obs}} - |\mathbf{F}_{PH}|^{\text{calc}} = |\mathbf{F}_{PH}|^{\text{obs}} - [|\mathbf{F}_P| \exp(i\alpha_P) + |\mathbf{F}_H| \exp(i\alpha_H)]$$

then for each heavy atom derivative j , and for each Bragg reflection (hkl) , a phase probability $P_j(\alpha_P)_{hkl}$ can be defined, making the assumption that all the errors reside in $|\mathbf{F}_{PH}|^{\text{calc}}$ and that errors follow a Gaussian distribution:

$$P_j(\alpha_P)_{hkl} = N \exp [-\varepsilon_j(\alpha_P)^2 / 2E_j^2]_{hkl}$$

where N is a normalization factor of the probability over 2π , and E_j is the error associated to the structure factor measurements. As expected, the distribution is bimodal if only one derivative is considered, the two peaks corresponding to the intersection points of the circles with radii $|\mathbf{F}_P|$ and $|\mathbf{F}_{PH}|$ (Fig. 4B and Fig. 5A). The total probability for each (hkl) reflection is obtained by multiplying the separate probabilities of each derivative j :

$$P(\alpha_P)_{hkl} = \prod_{j=1}^n P_j(\alpha_P)_{hkl}$$

The composed $P(\alpha_P)_{hkl}$ distribution normally shows a unimodal shape (Fig. 5B). An electron density map calculated with a weighted amplitude representing the centroid of the phase distribution yields the least error. The polar coordinates of the centroid define the phase angle $\alpha(\text{best})_{hkl}$ and a number m_{hkl} (varying between 0 and 1) called “figure of merit”, used as a weight associated to the corresponding native structure factor moduli $|\mathbf{F}_{hkl}|$:

$$\mathbf{F}_{hkl}(\text{best}) = m_{hkl} |\mathbf{F}_{hkl}| \exp [i\alpha_P(\text{best})_{hkl}]$$

The figure of merit corresponds to the cosine of the mean error in phase angle, providing an estimate of the quality of the phasing process. Reflections not properly phased maintain a bimodal distribution $P(\alpha_P)_{hkl}$, and a figure of merit close to 0. Viceversa, if the phasing has been accurate, then the probability distribution for each reflection $P(\alpha_P)_{hkl}$ is unimodal and m_{hkl} is close to 1. After the computation of the phase angles using the MIR method, it is possible to calculate an electron density map using native protein data and the phase angles derived from isomorphous replacement.

$$\rho(x,y,z) = 1/V \sum_{hkl} m_{hkl} |\mathbf{F}_{hkl}| \exp [i\alpha_P(\text{best})_{hkl}] \exp [-2\pi i(hx+ky+lz)]$$

Inspection of such electron density map allows to detect structural features typical of proteins (*i.e.* α -helices, β -strands, amino acid residues, etc), and to build an atomic model of the macromolecule. The knowledge of the atomic coordinates allows to calculate new phases that are usually more accurate of the “best” MIR phases. Then, the phase improvement will continue by the iteration of electron density map calculation, model rebuilding, and crystallographic refinement.

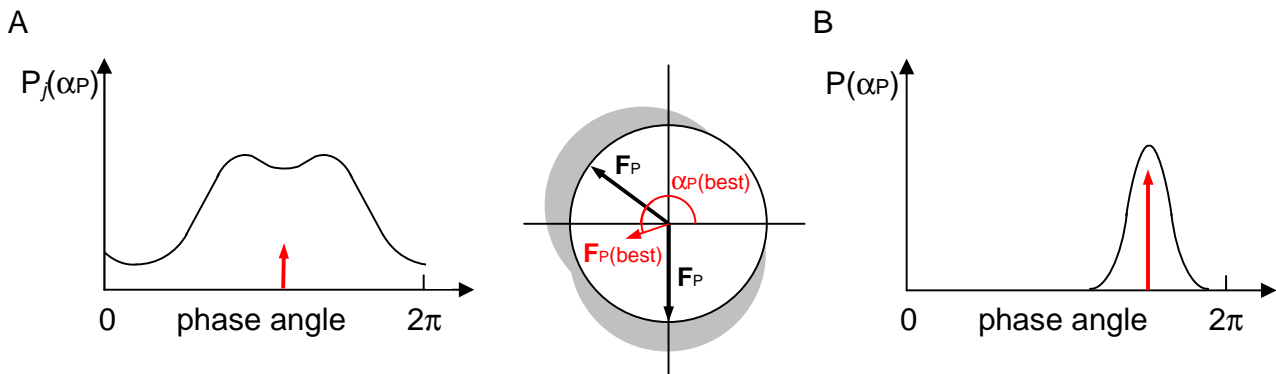


Figure 5. (A) Bimodal distribution, for the hkl reflection, of the probability $P_j(\alpha_P)$ for a phase angle α_P derived from derivative j . $F_P(\text{best})$ is the centroid of the distribution. (B) Unimodal distribution of the total probability $P(\alpha_P)$ for a phase angle α_P , derived from more than one derivative. (Adapted from [17])

2.4 SIRAS method

All elements, particularly those of high atomic number, display “anomalous scattering” when exposed to X-rays. The effect depends on the wavelength of the X-rays, being relevant when the energy of the incoming radiation is close to the element X-ray absorption edge. At such wavelength the energy of the X-ray photon is sufficient either to excite an electron of the strongly absorbing atom to a higher quantum state, or to eject the electron completely from the atom. As a consequence, the scattering factor f for the atomic species j (which is normally a real number for low atomic number elements) becomes “complex”, having three components:

$$f(\theta, \lambda)_j = f_0(\theta)_j + f'(\lambda)_j + if''(\lambda)_j$$

where $f_0(\theta)_j$ is the normal scattering term that is dependent on the Bragg angle θ , while $f'(\lambda)_j$ and $f''(\lambda)_j$ are the anomalous scattering dispersive and absorption terms which depend on the wavelength λ . The dispersive term (negative) decreases the normal scattering factor amplitude, whereas the absorption term is 90° advanced in phase and results in giving a complex nature to $f(\theta, \lambda)_j$. This leads to a breakdown in Friedel’s law, giving rise to small but significant anomalous differences between the absolute values and phases of $\mathbf{F}(hkl)$ and its Bijvoet mate $\mathbf{F}(-h-k-l)$. The anomalous or Bijvoet difference can be used in the same way as the isomorphous difference in Patterson method to locate the anomalous scatterers and can profitably be exploited for the determination of the protein phase angles. Figure 6A shows the origin of the differences between $\mathbf{F}(hkl)$ and $\mathbf{F}(-h-k-l)$ (Bijvoet pair).

The use of anomalous scattering in determining protein structures has increased recently in relation to the use of synchrotrons as X-ray sources [11, 12]. Synchrotron radiation is tunable, and its wavelength may, within certain limits, be tuned at will. This provides the opportunity of measuring the X-ray diffraction pattern of a protein crystal at wavelengths selected near (but also far) from the

absorption edges of any anomalous scatterer present (even as light as a sulfur atom). Phases for the native structure factors can then be derived in a similar way to the MIR case. Anomalous scattering can be used to break the phase ambiguity in a single isomorphous replacement experiment, leading to SIRAS (single isomorphous replacement with anomalous scattering) or MIRAS in case of multiple isomorphous replacement using anomalous scattering. Note that because of the 90° phase advance of the f'' term, anomalous scattering provides orthogonal phase information to the isomorphous term.

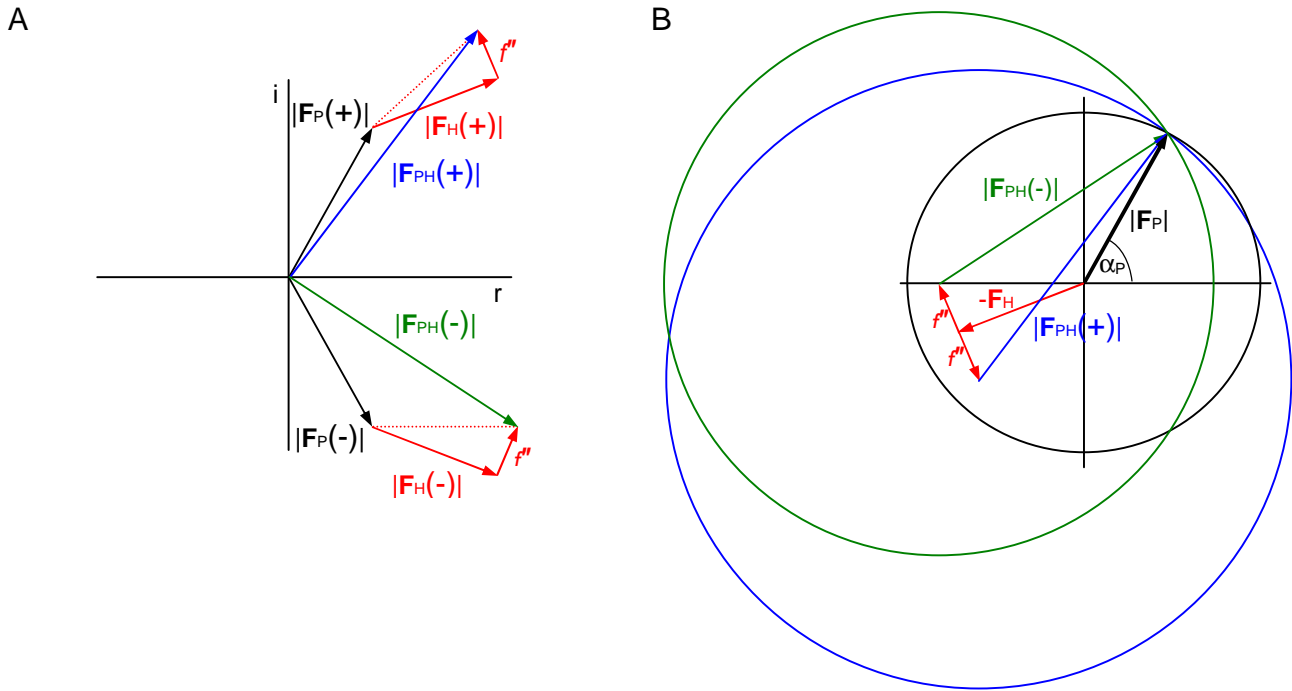


Figure 6. (A) Breakdown of Friedel's law if an anomalous scatterer is present: $\mathbf{F}(hkl) \neq \mathbf{F}(-h-k-l)$ or $\mathbf{F}_{PH}(+) \neq \mathbf{F}_{PH}(-)$. $\Delta F^\pm = |F_{PH}(+)| - |F_{PH}(-)|$ is the Bijvoet difference. (B) Harker construction for SIRAS.

2.5 MAD method

A different application of the anomalous scattering for macromolecular structure determination is the so called "Multiple wavelength Anomalous Dispersion" method (MAD). This method provides an extremely handy tool for the determination of protein phases, directly from just one crystal containing a good anomalous scatterer. The great advantage of using one crystal is that the lack of isomorphism problems are totally abolished. The anomalous scatterer may be provided by a conventional isomorphous derivative, by a selenium atom substituted to a sulfur atom in methionine and cysteine residues [18], or by an ion already present in native metalloproteins. In the case of engineered Se-substituted proteins the presence of one Se atom in a protein of approximately 150 amino acids is sufficient for a successful application of the MAD method [19].

MAD data are collected at several wavelengths, typically three, in order to maximize the absorption and dispersive effects. Typically, wavelengths are chosen at the absorption f'' peak (λ_1), at the point of inflection on the absorption curve (λ_2), where the dispersive term (which is the derivative of the f'' curve) has its minimum, and at a remote wavelength (λ_3 and/or λ_4). Figure 7 shows a typical absorption curve for an anomalous scatterer, together with the phase and Harker diagrams. The changes

in structure factor amplitudes arising from anomalous scattering are generally small and require accurate measurement of intensities. Importantly, the signal increases with resolution owing to the fall-off of normal scattering with resolution. The actual shape of the absorption curve must be determined experimentally by a fluorescence scan on the crystal at the synchrotron, as the environment of the anomalous scatterers can affect the details of the absorption. There is a need for excellent optics for accurate wavelength setting with minimum wavelength dispersion. Generally, all data are collected from a single frozen crystal to ensure the surviving of the same crystal to three data collections, with high redundancy in order to increase the statistical significance of the measurements and data are collected with as a high completeness as possible.

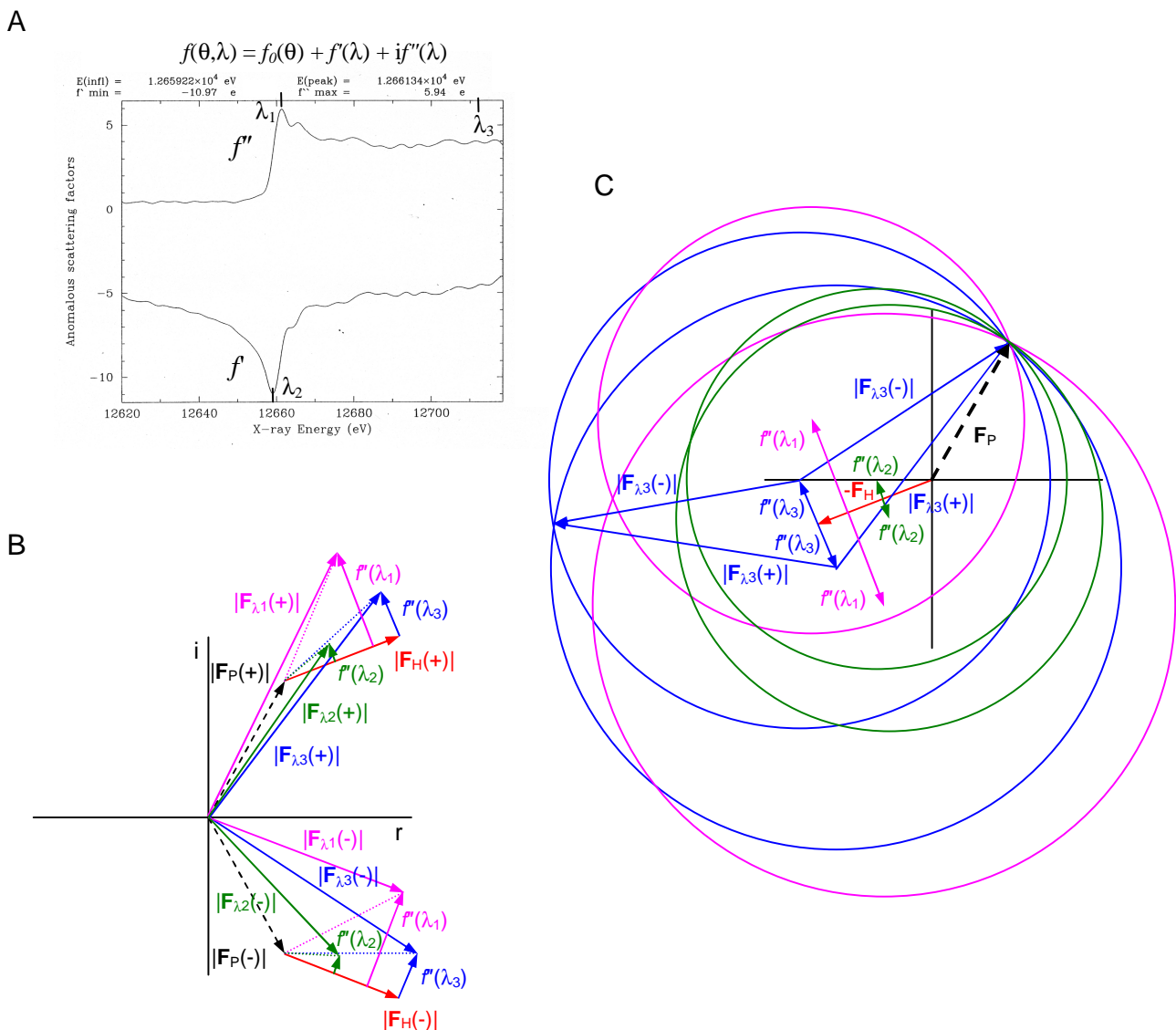


Figure 7. MAD phasing. (A) Typical absorption curve for an anomalous scatterer (in this example Se). (B) Phase diagram. $|F_P|$ is not measured (dashed line), so usually the remote λ_3 is chosen as the “native”. (C) Harker construction. (Adapted from [17])

2.6 Molecular replacement

The molecular replacement method is a phasing procedure of a native protein, that does not require specifically the use of synchrotron radiation. Here it is mentioned for completeness, since it represents one of the most common methods of protein phase determination in X-ray crystallography. The method is based on the observation that proteins which are homologous in their amino acid sequences have often a very similar three-dimensional fold. As a rule of thumb, a sequence identity >25% is normally required and an r.m.s. deviation of <2.0 Å between the Cα atoms of the model and the final new structure, although there are exception to this. Therefore, if one parent protein structure is available, it can be used to solve the 3D structure of homologous proteins, without the use of heavy atom derivatives. Conceptually, one has to transfer the known protein structure (search model) from its crystalline arrangement to the crystal of the protein of unknown structure (target protein). Such a procedure allows the calculation of approximated phases ($\alpha_{hkl}^{\text{calc}}$); these are combined with the observed structure factors ($|\mathbf{F}_{hkl}^{\text{obs}}|$) of the target protein, yielding a hybrid electron density map

$$\rho(x,y,z) = 1/V \sum_{hkl} |\mathbf{F}_{hkl}^{\text{obs}}| \exp(i\alpha_{hkl}^{\text{calc}}) \exp[-2\pi i(hx+ky+lz)]$$

which allows the iterative building of the target protein molecular model [13].

The correct positioning (orientation and translation) of the search model in the target unit cell is the key step for calculating starting phases sufficiently accurate to be able to model the target protein structure in the hybrid electron density map. Usually, the positioning of the search model in the target unit cell takes place in the Patterson space. If the model and the target protein are indeed similar and if they are oriented in the same way in unit cells of the same dimensions and symmetry, they should have very similar Patterson maps. We might imagine a trial-and error method in which we compute Patterson maps for several model orientations and compare them with the Patterson map of the target protein. In this manner, we could find the best orientation of the model, and then use that single orientation in the search for the best translational position of the model. In this two-step approach, the rotational and translational searches are defined by the functions:

$$C(\mathbf{R}) = \int_{\substack{\text{self} \\ \text{vectors}}} P_{\text{cryst}}(\mathbf{u}) P_{\text{mod}}(\mathbf{R},\mathbf{u}) d\mathbf{u}$$

$$D(\mathbf{t}) = \int_{\substack{\text{cross} \\ \text{vectors}}} P_{\text{cryst}}(\mathbf{u}) P_{1,2}(\mathbf{u},\mathbf{t}) d\mathbf{u}$$

where $P_{\text{cryst}}(\mathbf{u})$ and $P_{\text{mod}}(\mathbf{R},\mathbf{u})$ are the Patterson functions of the target protein crystal and of the search model rotated by matrix \mathbf{R} , respectively, and $P_{1,2}(\mathbf{u},\mathbf{t})$ is the cross-Patterson function of the model structure, properly translated by the vector \mathbf{t} within the unit cell. The rotational function $C(\mathbf{R})$ operates in a volume around the origin with a radius equal to the molecular diameter expected for the target protein. This region hosts all the intramolecular Patterson vectors (self-vectors set), which define the orientation of the search model with respect to the unit cell axes of the target crystal. The next step after the correct orientation of the search model is its proper translational positioning within the unit cell of the target crystal. The translational function $D(\mathbf{t})$ is calculated in the Patterson space, in the region external to the sphere used for the rotational search. This is because in that region lie the inter-molecular vectors (cross-vectors set) which define the position of the macromolecule relative to the unit cell symmetry elements.

If \mathbf{X} is the set of atomic (x,y,z) coordinates of the search model, \mathbf{R} is the rotational matrix obtained from the first step, and \mathbf{t} is the translational vector from the second step of the molecular replacement method, then

$$\mathbf{X}' = \mathbf{XR} + \mathbf{t}$$

provides the atomic coordinates of the search model in the target crystal [20]. It is from these coordinates that a set of preliminary phases can be calculated. These phases, combined with the structure factor amplitudes experimentally measured from the target crystal, produce the hybrid electron density map. Usually, more similar the structure of the search and target proteins are, less approximated is the calculated electron density map. The analysis of this map, using computer graphics, allows to build a starting molecular model of the target macromolecule. Such model will be subjected to several cycles of crystallographic refinement and modeling until the best model fitting the observed data is reached (see section 4).

3. Density modification

If the quality of the first electron density map is not good enough to allow a complete and unambiguous tracing of the polypeptide chain, improvement of the protein phase angles may be required prior to model building and refinement. Phase improvement can be based on two different techniques of “density modification” [13]. The first technique (known as “solvent flattening”) is based on the assumption that the protein molecules are characterised by regions of relatively high electron density whose boundaries can be located. On the contrary, the electron density map in the solvent regions, between protein molecules packed in the crystal lattice, is rather low. This is related to the dynamic nature of the solvent molecules in these regions of the unit cell, and results in the presence of noise peaks in the corresponding electron density. Such a noise can be removed by setting the electron density for the disordered solvent regions to a low constant value, related to its chemical composition. Then, new phases are calculated by back-transforming the modified electron density map. In the next step a new electron density map is calculated with experimentally observed structure factor amplitudes and with phase angles either from the solvent flattening alone or by combining them with phase angles obtained from isomorphous or molecular replacement method. The iteration of this procedure, along with the continuous updating of the envelope which separates the protein from the solvent region, and the addition (in small steps) of the data at higher resolution, is able to transform an initial hardly interpretable electron density map to a map where an atomic model can be easily built.

A similarly successful result can be achieved with the density modification technique known as “non-crystallographic symmetry averaging”. This technique is applied when more than one protein molecule is present in the crystal asymmetric unit. The assumption is that the electron density in molecules related by non-crystallographic symmetry (NCS) is essentially equal. If the NCS operators present in the asymmetric unit are known, then the electron density is averaged at each (x,y,z) point of NCS-related protein molecules, the solvent region flattened and the asymmetric unit is reconstituted. Phase angles from this new model are calculated by back-transforming the averaged electron density map. A new and improved density map is calculated with experimentally observed structure factor amplitudes and phase angles obtained from the NCS procedure [21]. If necessary, this phase information can be combined with previously known phase information. The result is improved signal-to-noise ratio and, in the end, a clearer image of the molecules. The NCS averaging method is spectacularly successful in systems with high symmetry, such as virus.

4. Refinement

4.1 Least squares refinement

The native protein phases, obtained by one of the methods discussed previously, allow the calculation of a preliminary electron density map, whose interpretation results in an approximated model of the protein structure showing the main features of the macromolecular architecture (Fig. 8). The partial model can be improved using the information present in the observed structure factor amplitudes. Such a process, known as crystallographic refinement, consists of iterative corrections of the model, to reach the best possible agreement between the calculated and observed structure factors. This procedure is essential to achieve the maximum reliability for the molecular model, as well as to provide the atomic resolution structural details needed for an effective biological research.

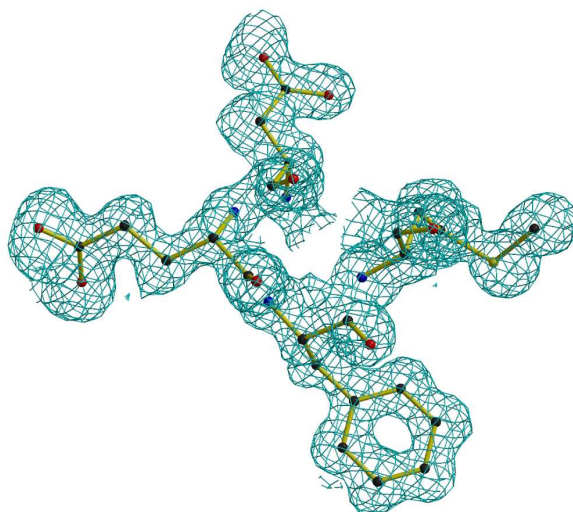


Figure 8. A small section of the electron density map (cyan mesh) contoured at 1σ , with the final structure superimposed (ball and stick representation).

During crystallographic refinement the molecular model is allowed to vary by changing the positional parameters \mathbf{p} , which are for each non-hydrogen atom j of the structure the coordinates (x_j, y_j, z_j) and the temperature factors (B_j). The temperature factors B_j are generally assumed to be isotropic, being related to the thermal oscillation (μ_j) of each atom j around its equilibrium position by the relation: $B_j = 8\pi^2 \langle \mu_j^2 \rangle$. The adjustment of the parameters (x_j, y_j, z_j, B_j) for all atoms within the initial approximated model is obtained by minimization of the function:

$$Q(\mathbf{p}) = \sum_{hkl} w_{hkl} (|\mathbf{F}_{hkl}|_{\text{obs}} - |\mathbf{F}(\mathbf{p})_{hkl}|_{\text{calc}})^2$$

The summation is calculated over all crystallographically independent reflections and w_{hkl} is the weight given to each observation ($w_{hkl} = 1/\sigma_{hkl}^2$, where σ_{hkl} is the standard deviation for the hkl observation). Least squares is the simplest statistical method used in macromolecular refinement. The least squares refinement of atomic coordinates and temperature factors is an iterative process. In each step the parameters to be refined shift only partially in the direction of their final values; usually, several refinement cycles are carried out before convergence is reached. The larger is the difference between

the observation and the model's prediction, the worse the model. The more precise is an observation, the more important that observation becomes in the overall sum. One varies the parameters of the model to find a set that gives the lowest sum of deviants.

Whereas for small molecules the ratio between observations (\mathbf{F}_{hkl}) and parameters to be refined is about 10:1, for macromolecules such a ratio is often close to 1. Such poor over-determination is generally compensated by the introduction of additional "observations", mostly related to the known stereochemical properties of the proteins (bond lengths, bond angles, van der Waals contacts, etc). The function which is then minimized is:

$$C = Q + kU$$

where Q is the crystallographic term previously introduced, U is the potential energy associated to a particular conformation of the macromolecule, and k is a scaling factor which controls the relative contribution of the energy and X-ray terms. U consists of several contributions based on the deviations of the stereochemical parameters of the protein model from ideal values, mainly derived from small molecule structures:

$$U = \sum_j 1/2 k_{b,j} (b_{j,calc} - b_{j,obs})^2 + \sum_j 1/2 k_{\tau,j} (\tau_{j,calc} - \tau_{j,obs})^2 + \sum_j k_{\phi} \{1 + \cos(n\phi_j - \delta)\} + \sum_{ij} [(Ar^{-12})_{ij} + (Br^{-6})_{ij}]$$

The potential energy function includes terms for bond stretching (b terms), bond angle bending (τ terms), torsion potentials, and van der Waals interactions, respectively. Electrostatic interactions are usually ignored, because they act over rather long distances and they are not particularly sensitive to small changes in atomic positions. When the crystallographic residual Q is minimized together with the potential energy term U, the refinement process is defined as "restrained" since the stereochemical parameters (treated as new "observations") are allowed to vary in a narrow range around their standard values. The atomic coordinates of the protein model vary within a sort of harmonic potential, driven by minimization of the crystallographic term. If the deviation of the model atoms from those of the real structure is large, the refinement may be trapped in a local minimum instead of converging to the correct solution. This situation can be avoided by correcting manually the model *via* computer graphics analysis, or by applying a more sophisticated refinement technique, which includes the contribution of molecular dynamics. This technique, which takes into account the kinetic energy associated to each atom of the macromolecule during the refinement, has rapidly gained popularity because of its wider radius of convergence which allows a faster refinement of the initial model [22].

4.2 Maximum-likelihood refinement

In general, the assumptions of the least squares method are that errors in the observations obey a normal distribution with completely known ("observed") variances and that, given perfect observations and the best parameters, the model would predict the observations perfectly. In many refinement problems, however, these assumptions are not completely correct. The simplest example occurs when the model is incomplete, say missing a domain. In this case it is impossible for any set of parameters to reproduce all the observations. The refinement function must account for the unknown contribution of the unmodeled part of the molecule and least squares cannot do that. Recently a more general approach has been introduced in the refinement procedure, by using the so called "maximum-likelihood" approach.

Maximum-likelihood is a general statistical framework for estimating the parameters of a model on the basis of observations [23]. This approach differs from least squares in that maximum-likelihood can accommodate observations with uncertainties of arbitrary character and model parameters whose values are also expected to have such uncertainties. While maximum likelihood is completely general, macromolecular refinement is such a difficult problem that no computer can perform a likelihood refinement in complete generality. Therefore the maximum likelihood method depends critically on the assumptions made about the nature of the uncertainties in the observations and the parameters of the final model [24]. The maximum-likelihood method begins with the assumption that the current structural model itself contains errors (Fig. 9A). Instead of a single location, as assumed by the least-squares method, there is a cloud of locations that each atom could occupy (usually, the distributions of positions are assumed to be normal and have equal standard deviations). This distribution of structures results in a distribution of values for the complex structure factors calculated from that model. Because of the nonlinear relationship between real and reciprocal space, the value of the structure factor calculated from the most probable model (labelled \mathbf{F}_{calc} in Fig. 9B) is, however, not the most probable value for the structure factor distribution. This is the key difference between least squares and maximum likelihood methods. The most probable value for the calculated structure factors has the same phase of \mathbf{F}_{calc} but a smaller amplitude which equalize \mathbf{F}_{calc} only when the model is infinitely precise. The width of the distribution, named σ_{calc} , also arises from the coordinate uncertainty and is large when the difference between \mathbf{F}_{calc} and the most probable value of the structure factor distribution is high. In refinement without experimental phase information, the probability distribution of the calculated values of the structure factors are converted to a probability distribution of the amplitude of this structure factor by integrating the two-dimensional distribution over all phase angles at each amplitude. This integral is represented by a series of concentric dashed circles in Fig. 9B. As expected, the most probable amplitude is smaller than $|\mathbf{F}_{\text{calc}}|$ (Fig. 9C).

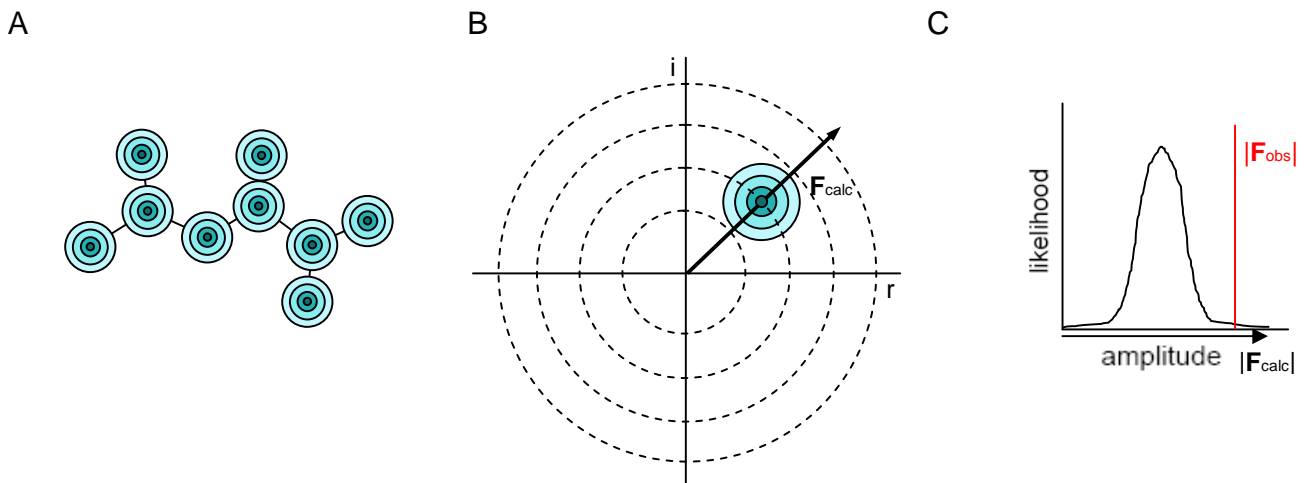


Figure 9. Probability distributions for one reflection in the maximum-likelihood approach (A) probability distributions of the atoms in the model. (B) The distribution of structures results in a distribution of values for the structure factors calculated from that model. \mathbf{F}_{calc} indicates the value of the structure factor calculated from the most probable model. A series of concentric dashed circles represent the integral over all phase angles needed to convert the probability distribution of the calculated value of the structure factor to a probability distribution of the amplitude of this structure factor. (C) The probability distribution for the amplitude of the structure factor. The arrow below the horizontal axis represents the amplitude of \mathbf{F}_{calc} , calculated from the most probable model. The corresponding $|\mathbf{F}_{\text{obs}}|$ is also indicated. (Adapted from[24])

With this distribution the likelihood of any value for $|\mathbf{F}|_{\text{obs}}$ can be evaluated, and one can calculate how to modify the model to increase the likelihood of $|\mathbf{F}|_{\text{obs}}$. For instance, in the case shown in (Fig. 9C) the likelihood of $|\mathbf{F}|_{\text{obs}}$ can be improved by either increasing $|\mathbf{F}|_{\text{calc}}$ or increasing the precision of the model. In general, compared to the least squares method, the basic maximum-likelihood residual function is:

$$Q(\mathbf{p}) = \sum_{hkl} (|\mathbf{F}_{hkl}|_{\text{obs}} - \langle |\mathbf{F}(\mathbf{p})_{hkl}|_{\text{calc}} \rangle)^2 / [\sigma_{\text{obs}}^2 + \sigma(\mathbf{p})_{\text{calc}}^2]$$

where $\langle |\mathbf{F}(\mathbf{p})_{hkl}|_{\text{calc}} \rangle$ is the expectation value of the amplitude of a structure factor calculated from all plausible model similar to \mathbf{p} , and $\sigma(\mathbf{p})_{\text{calc}}$ is the width of the distribution of values for $|\mathbf{F}(\mathbf{p})_{hkl}|_{\text{calc}}$.

4.3 Quality of the model

The progress of the crystallographic refinement is monitored by comparing the measured structure factor amplitudes ($|\mathbf{F}_{hkl}|_{\text{obs}}$) with those calculated from the current atomic model ($|\mathbf{F}_{hkl}|_{\text{calc}}$). As the model converges to the correct structure, the calculated structure factor amplitudes should also converge to the corresponding measured values. The most widely used measure of convergence is the so called “residual index”, or R -factor:

$$R = \frac{\sum_{hkl} ||\mathbf{F}_{hkl}|_{\text{obs}} - |\mathbf{F}_{hkl}|_{\text{calc}}|}{\sum_{hkl} |\mathbf{F}_{hkl}|_{\text{obs}}}$$

Values of the R -factor range from zero, for perfect agreement of calculated and observed structure factors, to about 0.6, when the measured structure factors are compared to a set of random amplitudes. An early model with R -factor near 0.4 is promising and is likely to improve during the refinement procedure. A desirable target R -factor for a refined protein model with data to 2.5 Å is about 0.2.

A more revealing criterion of model quality and improvement is the so called “free R -factor”, or R_{free} . The R_{free} is calculated exactly as the R -factor but the summations are on a small set of randomly chosen reflections (5-10% of the total measured reflections), the “test set”, which are set aside from the beginning and never used during the refinement procedure. Therefore, at any stage in refinement, R_{free} measures how well the current atomic model predicts a subset of measured amplitudes that were not included in the refinement of the model. For this reason, R_{free} provides a less biased measure of the overall model quality, correlating very well with phase accuracy of the atomic model. In general, during intermediate stages of refinement, R_{free} values are higher than the R -factor, but in the final stages, the two often become more similar (the final R_{free} value is generally higher than the R -factor by 2 - 7%).

In addition to R -factors as indicators of convergence, other structural parameters can be monitored to check whether the model is chemically, stereochemically, and conformationally reasonable. During the progress of refinement, all of these structural parameters should improve.

5. Data bases

The macromolecular structures solved by means of X-ray crystallography or NMR techniques are usually submitted to the Protein Data Bank (<http://www.rcsb.org/pdb>) as text files [25]. The head of each PDB file contains specific information about the macromolecule, whereas the body of the file consists of a list of coordinates for each single atom of the structure (expressed in Å and referred to

orthogonal axes), with their temperature factors. Additionally, the atomic coordinates of a variable number of experimentally located solvent molecules (usually water molecules) are also present in the PDB files. These solvent molecules are localized on the surface of the macromolecule as well as in internal cavities. A substantial revision of the PDB is currently in progress with the aim of improving the quality and the completeness of the deposited structural and biological information. For instance, the PDB provides several links to correlated data bases and access to structural analysis software for the study of protein folds, general research in various fields of structural biology, and for the validation of each new deposition. As a rule of thumb, the basic criteria that an occasional user should take into account to check the quality of a deposited structure are:

- (a) the agreement index between calculated and observed structure factors (R -factor and R_{free})
- (b) maximum resolution of the data used during the refinement procedure
- (c) global stereochemistry of the macromolecule (as evaluated by estimates of the deviation from the ideal values for bond length, bond angles, etc)
- (d) atomic temperature factors (average and local values) of the macromolecule and of the localized solvent molecules

The amount of data provided by bio-crystallographic analysis is continuously increasing. During the year 2010, on average, one new structure has been deposited in the PDB every 1.5 hours. This flood of information, however, does not necessarily correspond to novel structural information, since many protein folds are well conserved through the different evolutionary *phyla*. Statistics and biological considerations suggest that the all proteins present in Nature (about 10^5 different types in humans) are compatible with the existence of a limited number of folds, probably less than 10^4 .

6. Structural genomics

Macromolecular crystallography is experiencing a surge in new methodologies, largely fuelled by developments in structural genomics and by the needs of pharmaceutical industry. While most structural biologists pursue structures of individual proteins or protein groups, specialists in structural genomics pursue structures of proteins on a genome wide scale, in particular human proteins implicated in disease states, as well as proteins from a set of pathogens. The Human Genome Project has produced an enormous body of information about the sequence of the human genome and in addition genome sequences of a large number of human pathogens, from tuberculosis to malaria, are also known. However, detailed information on the structure, function and interaction of the tens of thousands of proteins encoded by these genomes is required in order to fully exploit this new panoply of data.

Rapidly increasing the numbers of known atomic structures of proteins will have a number of strategic benefits for medical research, basic and applied. In particular, knowledge of the three-dimensional structures of key human proteins or human pathogenic proteins is vital for speeding up the difficult task of discovering new antibiotics or anti-cancer drugs (Center for Structural Genomics of Infectious Diseases: <http://csgid.org>). X-ray crystallography at latest generation synchrotron sources is the key technique that will build the database of three-dimensional information on protein structure and by generating large amounts of systematic crystallization data that may lead to more predication-based crystallization methods. Nuclear magnetic resonance techniques and neutron scattering constitute further important and complementary structure determination tools.

As opposed to traditional structural biology, the determination of a protein structure through a structural genomics effort often (but not always) comes before anything is known regarding protein function. This implies new challenges in structural bioinformatics, *i.e.* determining protein function from its three-dimensional structure and the need to speed up the structure determination process by optimising each step for high throughput, including target selection, protein production, purification,

crystallization and structure analysis by X-ray crystallography and/or NMR spectroscopy. As an example, several synchrotron beamlines normally provide the so called “remote data collection system” for macromolecular crystallography, which allows users to work on their samples, shipped and mounted on the beamline goniometer, through the use of extensive robotics that allow for comprehensive crystal manipulation, screening, selection of best samples, and data collection, all done from their home laboratory. The user can also conduct data processing and phasing calculations during the actual data collection experiment (*i.e.* “on the fly”). This service not only minimizes the cost of data collection, but also simplifies the operation of the beamline, thereby reducing human traffic and security issues.

References

- [1] Branden C. and Tooze J., *Introduction to Protein Structure* (2nd Ed., Garland Publishing, Inc, New York) 1999.
- [2] Matthews B.W., *Methods in Enzymology*, **276** (1997) 3-10.
- [3] Gane P.G. & Dean P.M., *Curr. Opin. Struct. Biol.*, **10** (2000) 401-404.
- [4] Erlandsen H., Abola E.E. and Stevens R.C., *Curr. Opin. Struct. Biol.*, **10** (2000) 719-730.
- [5] Ban N., Nissen P., Hansen J., Moore P.B. and Steitz T.A., *Science*, **289** (2000) 905-920.
- [6] Ducruix A. and Giegé R., *Crystallization of nucleic acids and proteins: a practical approach* (Oxford University Press) 1992.
- [7] McPherson A., *Crystallization of Biological Macromolecules* (Cold Spring Harbor Laboratory Press) 1999.
- [8] Weber P.C., *Methods in Enzymology*, **276** (1997) 13-22.
- [9] Chayen N.E. and Helliwell J.R., *Ann. N.Y. Acad. Sci.*, **974** (2002) 591-597.
- [10] Ravelli R.B.G. and Garman E.F., *Curr. Opin. Struct. Biol.*, **16** (2006) 624-629.
- [11] Helliwell J.R., *Macromolecular Crystallography with Synchrotron Radiation* (Cambridge University Press) 1992.
- [12] Helliwell J.R., *Methods in Enzymology*, **276** (1997) 203-217.
- [13] Drenth J., *Principles of Protein Crystallography* (2nd Ed., Springer-Verlag) 1999.
- [14] Bourgeois D., Wagner U. and Wulff M., *Acta. Crystallogr. D*, **56** (2000) 973-985.
- [15] Schotte F., Soman J., Olson J.S., Wulff M. and Anfinrud P.A., *J. Struct. Biol.*, **147** (2004) 235-246.
- [16] Milani M., Nardini M., Pesce A., Mastrangelo E. and Bolognesi M., *IUBMB Life*, **60** (2008) 154-158.
- [17] Taylor G., *Acta. Crystallogr. D*, **59** (2003) 1881-1890.
- [18] Doublé S., *Methods in Enzymology*, **276** (1997) 523-530.
- [19] Hendrickson W.A. and Ogata C.M., *Methods in Enzymology*, **276** (1997) 494-523.
- [20] Navaza J. and Saludjian P., *Methods in Enzymology*, **276** (1997) 581-594.
- [21] Vellieux F.M.D. and Read R.J., *Methods in Enzymology*, **277** (1997) 18-53.
- [22] Brünger A.T. and Rice L.M., *Methods in Enzymology*, **277** (1997) 243-269.
- [23] Pannu N.S. and Read R.J., *Acta. Crystallogr. A*, **52** (1996) 659-668.
- [24] Tronrud D., *Acta. Crystallogr. D*, **60** (2004) 2156-2168.
- [25] Berman H.M., Westbrook J., Feng Z., Gilliland G., Bhat T.N., Weissig H., Shindyalov I.N. and Bourne P.E., *Nucleic Acids Res.*, **28** (2000) 235-242.