Published in final edited form as:

Methods Mol Biol. 2017 ; 1607: 349–356. doi:10.1007/978-1-4939-7000-1_14.

Many ways to derivatize macromolecules and their crystals for phasing

Mirosława Dauter¹ and Zbigniew Dauter²

¹Leidos Biomedical Research, Inc., Basic Science Program, Argonne National Laboratory, Argonne, IL 60439, USA

²Synchrotron Radiation Research Section, MCL, National Cancer Institute, Argonne National Laboratory, Argonne, IL 60439, USA

Abstract

Due to the availability of many macromolecular models in the Protein Data Bank, the majority of crystal structures are currently solved by molecular replacement. However, truly novel structures can only be solved by one of the versions of the special-atom method. The special atoms such as sulfur, phosphorus or metals could be naturally present in the macromolecules, or could be intentionally introduced in a derivatization process. The isomorphous and/or anomalous scattering of X-rays by these special atoms is then utilized for phasing. There are many ways to obtain potentially useful derivatives, ranging from the introduction of special atoms to proteins or nucleic acids by genetic engineering or by chemical synthesis, to soaking native crystals in solutions of appropriate compounds with heavy and/or anomalously scattering atoms. No approach guarantees the ultimate success and derivatization remains largely a trial-and-error process. In practice, however, there is a very good chance that one of a wide variety of the available procedures will lead to successful structure solution.

Keywords

Derivatization of crystals; heavy-atoms; anomalous signal; MIR phasing; MAD phasing; SAD phasing

1. Introduction

There are two principal methods of solving macromolecular crystal structures from diffraction data by the technique of macromolecular crystallography (MX). If a model sufficiently similar to the investigated macromolecule is available, the unknown structure may be determined by the method of molecular replacement. Otherwise, the only practically available approach is to use one of the versions of the approach that can be termed the “special atom method”. Such a procedure utilizes some special properties of a small number of certain atoms present among a large number of “standard” elements (C, N, O, H) within the macromolecule. The first step in this approach is the location of the special atoms, and the next stage extends this special-atom “substructure” to the complete crystal structure. The special characteristics may be a large number of electrons of some heavy atoms, anomalous X-ray scattering properties of both the heavier or lighter atoms, or a combination of both.

The heavy and/or anomalous atoms may be naturally present in the crystallized molecules (e.g. in metalloproteins), otherwise they have to be incorporated into native macromolecules before or after crystallization in a process known as derivatization.

Currently there are ~120,000 structures in the Protein Data Bank [1] that can be used as potential search models for molecular replacement and, indeed, the majority of crystal structures are nowadays solved by this technique. However, if no suitable model can be found, there is a need to resort to the special-atom approach. This method was used for solving the first X-ray structures of hemoglobin [2, 3], myoglobin [4] and in other pioneering works. A very illuminating account of the process of crystal structure determination of lysozyme in the early 1960s was presented by Blake et al. [5]. In the early days, when diffraction data were recorded on photographic films, their accuracy was only sufficient to utilize the isomorphous signal of heavy-atom derivatives, with differences between reflection intensities of the native and derivative crystals amounting sometimes to as much as 15–25%. Only after the introduction of more accurate ways of measuring reflection intensities (using multiwire, imaging plate, CCD, and pixel detectors), and with the advent of powerful and tunable synchrotron X-ray beam lines, has it become possible to achieve diffraction data accuracy (on the order of a few percent), necessary for productive exploitation of the inherently weak anomalous signal of various elements present in macromolecular crystals. Currently the anomalous signal is used to solve a great majority of novel X-ray crystal structures, mostly thanks to the efforts of Wayne Hendrickson, who pioneered the methods of multiwavelength anomalous diffraction (MAD [6]), as well as single-wavelength anomalous diffraction (SAD [7]).

Thus, to solve a novel X-ray crystal structure of a protein or nucleic acid, one has to utilize the signal, isomorphous or anomalous, of some special atoms present in the investigated structure. These atoms might be present in the native molecules, for example, in various metalloproteins containing such metals as Fe, Cu, Zn, etc., or could be even lighter elements, such as sulfur contained in almost all proteins and phosphorus present in all nucleic acids. The suitable elements can be introduced by genetic engineering, as is the case of selenium incorporated in the form of selenomethionine (SeMet) [8], which nowadays is the workhorse of protein crystallography owing to its relatively strong anomalous signal. Of course, the classic approach is based on the introduction of heavy metals, such as Pt, Hg, Au, Os, lanthanides etc., by soaking crystals in buffers also containing appropriately selected compounds. The metal cations are coordinated by reactive functional groups at the protein surface, such as the sulfhydryl group of cysteines, nitrogen atoms of histidines, or carbonyl and carboxylate oxygen atoms. Very useful for phasing large crystal structures are multi-center metal complexes [9, 10] which provide large phasing signals, especially at low resolution. Some ions, such as halides, Br⁻ and I⁻, do not form covalent or coordination bonds with the proteins, but stick to their surface by hydrogen bonds or by non-polar interactions, often sharing their locations with solvent water molecules [11].

There is, therefore, a large palette of approaches which can be used to solve novel X-ray crystal structures of proteins and nucleic acids when the application of molecular replacement is not possible or not successful. In general, the experimental data used for special-atom phasing have to be more accurate than those for molecular replacement.

Phasing based on the anomalous signal of sulfur or phosphorus requires exceptionally accurate data, since the expected phasing signal may be at the level of 1% or less [12–14]. Whereas data used for structure refinement can tolerate a certain degree of radiation damage, those used for the isomorphous and especially anomalous applications should be collected so as to avoid at all cost any radiation damage [15]. However, radiation damage can sometimes be used for phasing, in a method known as RIP [16]. It is possible to combine data collected from several crystals [17], provided they are isomorphous.

2. Incorporation of special atoms

The special atoms, intended as the source of the isomorphous and/or anomalous phasing signal, can be incorporated into macromolecular crystals in a variety of ways. Obviously, to exploit the signal of sulfur, which is inherently present in most proteins, or of transition metals found in metalloproteins, no additional derivatization is necessary. If the investigated macromolecule does not contain any elements suitable for phasing, it is necessary to introduce them before conducting the diffraction experiment. This can be done in several ways.

2.1. Modification of the macromolecules before crystallization

The first possibility is to prepare by chemical or biochemical methods a variant of the protein or nucleic acid containing the special atom in advance of the crystallization trials. Incorporation of selenomethionine by genetic engineering [8] is a very commonly used way of obtaining a convenient vehicle for solving protein crystal structures by the MAD or, more frequently, SAD approaches. Indeed, this is the way how the majority of novel protein crystal structures are solved these days. It is also possible to treat other elements in a similar way, e.g. by incorporating in the protein sequence the unnatural amino acid p-iodophenylalanine instead of phenylalanine [18]. The iodine atom can be substituted at the aromatic rings of tyrosine by treating the protein with N-iodosuccinimide prior to crystallization [19]. Analogously to SeMet in proteins, 5-bromouracil can be used as an anomalous marker in chemically synthesized nucleic acids, also in their complexes with proteins [20]. It is also possible to introduce selenium into nucleic acids in the form of phosphoroselenates [21]. Another possibility is to utilize the RIP or “Cheshire cat” effect of isomorphous differences originating from radiation-induced disappearance of the heavy atoms, as demonstrated by the solution of the crystal structures of an Hg-derivative of a protein [22] and bromouracil-modified nucleic acid [23].

2.2 Classic heavy-atom derivatization

The standard way of obtaining heavy-atom derivatives, introduced to protein crystallography as the first method of phasing macromolecular crystal structures by Perutz [3] and his followers [4, 5], is based on soaking native crystals in aqueous solutions containing salts of the appropriate metals. In a variant of this method, the protein or nucleic acid is crystallized from a solution containing the selected salt. A variety of different inorganic and organic salts have been utilized [24], but the most popular and successful reagents are the so-called magic seven: $\text{K}_2\text{PtCl}_4\text{KAu}(\text{CN})_2\text{K}_2\text{HgI}_4\text{UO}_2(\text{AcO})_2\text{HgCl}_2\text{K}_3\text{UO}_2\text{F}_5$, and PCMBs (para-chloro mercury benzoic acid sulfonate) [25]. A compendium of various reagents and derivatives is

available at the Heavy Atom Databank (<http://www.sbg.bio.ic.ac.uk/had/> [26]). An important factor to keep in mind is the high toxicity of many heavy-metal compounds, especially those containing mercury or osmium, and appropriate safety procedures must be strictly observed when working with such reagents.

The ligands (such as Cl^- , Br^- , I^- , CN^- , etc.) in the coordination compounds of the heavy metals must hydrolyze or be substituted by chemically reactive groups of proteins for their successful derivatization. This process can be rapid (seconds), or might take a considerable time (months). Moreover, some reactions with heavy metals may sometimes induce structural rearrangements of proteins, introducing significant non-isomorphism, or even cause visible cracking of the soaked crystals. The usual practice is to soak protein crystals in diluted, millimolar solutions of heavy-atom salts for longer time (several hours or days). If the crystals visibly deteriorate when observed under a microscope, the concentration of the heavy-atom reagent should be lowered. In fact, such a behavior has a positive side, providing a confirmation of a successful, even if too vigorous, derivatization. The excess of the unbound heavy-atoms, which is not productive for phasing, may be removed by a short soak in the mother liquor devoid of derivatization agents. This is particularly important for highly absorbing salts of very heavy metals, such as osmium. On the other hand, care must be taken not to back-soak the metal from the productive protein-binding sites.

As a modification of the usual, slow heavy-atom soaking procedure, the quick soaking approach was proposed [27], where the heavy atoms bind to the protein rapidly, before they can introduce any significant non-isomorphism or structural rearrangements of the crystal contents.

The choice of the most promising derivative is difficult and often the only way is to try a number of reagents, hoping for the eventual success, which may depend on many factors, such as the contents of the crystallization buffer, its pH, concentration, temperature, etc. If the protein contains free sulfhydryl groups, not engaged in disulfide bridges, a promising approach is to attempt mercury derivatization, using one of the many Hg reagents available, such as $\text{HgCl}_2\text{K}_2\text{HgI}_4\text{Hg}(\text{CH}_3)\text{Cl}$, PCMBs, thiomersal [EMTS, ethylmercury thiosalicylate], mersalyl ($\{3\text{-[2-(carboxymethoxy)benzoyl]amino-2-methoxypropyl}\}$ [hydroxyl] mercury), and others. Compounds containing Pt, Au, Os and similar elements are often coordinated by the nitrogen atoms of the imidazole rings of histidine or the sulfur atom of methionine. Soaking of native protein crystals in solutions containing triiodides (I_3^-) may lead to iodination of tyrosine rings [28].

A separate group of reagents, especially useful for phasing very large structures of proteins and complexes, are multinuclear metal clusters, such as $\text{Ta}_6\text{Br}_{12}^{2-}$ or $\text{P}_2\text{W}_{18}\text{O}_{62}^{6-}$ [10]. They were very helpful in cracking the structure of the ribosome [9, 29, 30].

The success of derivatization becomes apparent only after collecting the diffraction data and comparing them with the native set or, in fact, after a successful structure solution. However, some other symptoms may provide useful indications also at earlier stages. For example, the mass spectra of potentially derivatized macromolecules may confirm the successful binding of heavy atoms [25, 31]. Some heavy-atom compounds distinctly colorize the transparent

crystals; for example the “magic green” tantalum bromide complex $\text{Ta}_6\text{Br}_{12}^{2-}$ makes the crystals dark green after successful binding [32]. Often the whole green color is “soaked” from the crystallization solution into crystals, which indicates that the soaking drop should be supplemented with a fresh dose of the reagent.

2.3. Quick halide soaks

In contrast to heavy-atom reagents, the bromide and iodide ions do not form stable bonds with proteins and they penetrate into crystals very rapidly, even during a few second soaks [13]. They populate multiple sites around the protein surface, forming ion pairs with the positively charged side chains of Arg and Lys; hydrogen bonds with the amide or hydroxyl donors; or sit in hydrophobic niches. In a variant of this approach it is possible to use triiodides I_3^- , easily prepared by dissolving elemental iodine in the solution of KI [33].

Usually a large number of partially occupied Br^- or I^- sites can be identified as they share their sites with water molecules. The Br^- ions are suitable for SAD or MAD phasing since the $K\alpha$ X-ray absorption edge of Br is at 0.92 Å, easily achievable at all synchrotron beam lines. The absorption edges of I^- iodine are not easily accessible, but iodine has a significant anomalous effect, especially at wavelengths longer than 1.5 Å, and is very convenient for SAD phasing using copper radiation data collected at home laboratories.

The halide concentration in the soaking solution should be high, up to 1 M or more, although some successful results were obtained at concentrations lower than 0.2 M. It is advisable to start testing with high concentration of NaBr or NaI and observe if the crystal survives such a treatment without visible cracking or dissolving. If a high concentration of the halides quickly deteriorates the crystal quality, it should be lowered and the procedure repeated. Due to the fast diffusion of halides into protein crystals, in practice it is enough to sweep the crystal through a drop of the cryoprotecting buffer supplemented with the halides immediately prior to freezing it for data collection.

2.4. Incorporation of noble gases

Noble gases, such xenon or krypton, are capable of penetrating into protein crystals under increased pressure and occupy sites at hydrophobic patches on the protein surface [34]. In practice, crystals mounted in nylon loops or in capillaries are kept in high-pressure cells for up to 1 hour under a noble gas pressure of several MPa and then rapidly flash-cooled in cold nitrogen gas or liquid. Since Xe or Kr are inert and do not react with proteins chemically, they usually do not introduce significant non-isomorphism. The atoms of Xe and Kr are isoelectronic with, respectively, the I^- and Br^- ions and have analogous X-ray scattering properties. Thus, xenon derivatives display significant isomorphous and anomalous signals at longer wavelengths and krypton is suitable for MAD phasing. Pressure cells of different construction for noble gases derivatization are available commercially and can be found at many synchrotron beam lines.

3. Conclusions

A large palette of techniques exists for obtaining useful derivatives of macromolecular crystals with the incorporation of a wide selection of special atoms suitable for phasing by

the MIR, MIRAS, MAD or SAD methods. However, no technique guarantees a successful structure solution. Currently, the most popular and most successful is phasing of novel crystal structures using the anomalous signal of Se, introduced to proteins as SeMet by genetic engineering, but even this universal approach is not always applicable.

References

1. Berman HM, Westbrook J, Feng Z, et al. The Protein Data Bank. *Nucleic Acids Res.* 2000; 28:235–242. [PubMed: 10592235]
2. Green DW, Ingram VM, Perutz MF. The structure of haemoglobin. IV. Sign determination by the isomorphous replacement method. *Proc R Soc Lond.* 1954; 225:287–307.
3. Perutz MF, Rossmann MG, Cullis AF, et al. Structure of haemoglobin: a three-dimensional Fourier synthesis at 5.5-Å resolution, obtained by X-ray analysis. *Nature.* 1960; 185:416–421. [PubMed: 18990801]
4. Kendrew JC, Bodo G, Dintzis HM, et al. A three-dimensional model of the myoglobin molecule obtained by X-ray analysis. *Nature.* 1958; 181:662–666. [PubMed: 13517261]
5. Blake, CCF., Fenn, RH., Johnson, LN., et al. *International Tables for Crystallography. Vol. F.* Kluwer Academic Publishers; Dordrecht: 2001. A historical perspective: how the structure of lysozyme was actually determined; p. 745–772.
6. Hendrickson WA, Teeter MM. Structure of the hydrophobic protein crambin determined directly from the anomalous scattering of sulfur. *Nature.* 1981; 290:107–113. [PubMed: 28769131]
7. Hendrickson WA. Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science.* 1991; 254:51–58. [PubMed: 1925561]
8. Hendrickson WA, Horton JR, LeMaster DM. Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three dimensional structure. *EMBO J.* 1990; 9:1665–1672. [PubMed: 2184035]
9. Gluehmann M, Zarivach R, Bashan A, et al. Ribosomal crystallography: From poorly diffracting microcrystals to high-resolution structures. *Methods.* 2001; 25:292. [PubMed: 11860283]
10. Dauter Z, Dauter M, Rajashankar KR. Novel approach to phasing proteins: derivatization by short cryo-soaking with halides. *Acta Crystallogr.* 2000; D56:232–237.
11. Dauter Z. Use of polynuclear metal clusters in protein crystallography. *Compt Rend Chimie.* 2005; 8:1808–181.
12. Wang BC. Resolution of phase ambiguity in macromolecular crystallography. *Methods Enzymol.* 1985; 115:90–112. [PubMed: 4079800]
13. Ramagopal UA, Dauter M, Dauter Z. Phasing on anomalous signal of sulfurs: what is the limit? *Acta Crystallogr.* 2003; D59:1020–1027.
14. Dauter Z, Adamiak DA. Anomalous signal of phosphorus used for phasing DNA oligomer: importance of data redundancy. *Acta Crystallogr.* 2001; D57:990–995.
15. Garman EF. Radiation damage in macromolecular crystallography: what is it and why should we care? *Acta Crystallogr.* 2010; D66:339–351.
16. Ravelli RBG, Leiros HK, Pan B, et al. Specific radiation damage can be used to solve macromolecular crystal structures. *Structure.* 2003; 11:217–224. [PubMed: 12575941]
17. Liu Q, Liu Q, Hendrickson WA. Robust structural analysis of native biological macromolecules from multi-crystal anomalous diffraction data. *Acta Crystallogr.* 2013; D69:1314–1332.
18. Xie J, Wang L, Brock A, et al. The site-specific incorporation of p-iodo-L-phenylalanine into proteins for structure determination. *Nature Biotechnol.* 2004; 22:1297–1301. [PubMed: 15378068]
19. Brzozowski AM, Derewenda U, Derewenda ZS, et al. A model for interfacial activation in lipases from the structure of a fungal lipase-inhibitor complex. *Nature.* 1991; 351:491–494. [PubMed: 2046751]
20. Hendrickson WA, Ogata CM. Phase determination from multiwavelength anomalous diffraction measurements. *Methods Enzymol.* 1997; 276:494–513. [PubMed: 27799111]

21. Wilds CJ, Pattanayek R, Pan C, et al. Selenium-assisted nucleic acid crystallography: use of phosphoroselenates for MAD phasing of a DNA structure. *J Am Chem Soc.* 2002; 124:14910–14916. [PubMed: 12475332]
22. Ramagopal UA, Dauter Z, Thirumuruhan R, et al. Radiation-induced site-specific damage of mercury derivatives: phasing and implications. *Acta Crystallogr.* 2005; D61:1289–1298.
23. Ennifar E, Carpentier P, Ferrer JL, et al. X-ray induced debromination of nucleic acids at the Br K absorption edge and implications for MAD phasing. *Acta Crystallogr.* 2002; D58:1262–1268.
24. Agniswamy J, Joyce MG, Hammer CH, et al. Towards a rational approach for heavy-atom derivative screening in protein crystallography. *Acta Crystallogr.* 2008; D64:354–367.
25. Boggon TJ, Shapiro L. Screening for phasing atoms in protein crystallography. *Structure.* 2000; 8:R143–R149. [PubMed: 10903954]
26. Carvin D, Islam SA, Sternberg MJE, et al. A databank of heavy-atom binding sites in protein crystals: a resource in use for multiple isomorphous replacement and anomalous scattering. *Acta Crystallogr.* 1998; D54:1199–1206.
27. Sun PD, Radaev S, Kattah M. Generating isomorphous heavy-atom derivatives by a quick-soak method. Part. I. test cases. *Acta Crystallogr.* 2002; D58:1092–1098.
28. Kretsinger RH. A crystallographic study of iodinated sperm whale metmyoglobin. *J Mol Biol.* 1968; 31:315–318. [PubMed: 5635535]
29. Ban N, Freeborn B, Nissen P, et al. A 9 Å resolution X-ray crystallographic map of the large ribosomal subunit. *Cell.* 1998; 93:1105–1115. [PubMed: 9657144]
30. Clemons WM, May JLC, Wimberly BT, et al. Structure of a bacterial 30S ribosomal subunit at 5.5 Å resolution. *Nature.* 1999; 400:833–840. [PubMed: 10476960]
31. Joyce MG, Radaev S, Sun PD. A rational approach to heavy-atom derivative screening. *Acta Crystallogr.* 2010; D66:358–366.
32. Pasternak O, Bujacz A, Biesiadka J, et al. MAD phasing using the $(\text{Ta}_6\text{Br}_{12})^{2+}$ cluster: a retrospective study. *Acta Crystallogr.* 2008; D64:595–606.
33. Evans G, Bricogne G. Triiodide derivatization in protein crystallography. *Acta Crystallogr.* 2003; D59:1923–1929.
34. Prangé T, Schiltz M, Pernot L, et al. Exploring hydrophobic sites in proteins with xenon or krypton. *Proteins.* 1998; 30:61–73. [PubMed: 9443341]