International Tables for Crystallography (2006). Vol. F, Section 12.1.2, pp. 247–248.

12. ISOMORPHOUS REPLACEMENT

12.1. The preparation of heavy-atom derivatives of protein crystals for use in multiple isomorphous replacement and anomalous scattering

BY D. CARVIN, S. A. ISLAM, M. J. E. STERNBERG AND T. L. BLUNDELL

12.1.1. Introduction

The traditional method of multiple isomorphous replacement (MIR) was introduced by Perutz and co-workers in 1954 (Green *et al.*, 1954) and is often enhanced by anomalous scattering (MIRAS) [see Blundell & Johnson (1976) for a review]. The method remains popular for solution of the phase problem in the absence of the structure of a close homologue, although the use of multiple anomalous dispersion is likely to increase in the coming years (Hendrickson, 1985).

Protein crystals comprise an open lattice of protein molecules with solvent occupying the channels and spaces which normally comprise between 30 and 80% of the crystal volume. The preparation of a useful derivative requires the binding of a heavy atom to a specific position, usually on the protein surface, for example by the displacement of a lighter solvent molecule or an ion, without distorting the protein or crystal lattice.

Ideally, rational selection of suitable heavy-atom reagents requires a comprehensive knowledge and understanding of the crystalline structure of the protein. Normally, this information is unavailable since it is the objective of the crystal structure analysis! Nevertheless, the sequence and mechanism of action may suggest which heavy-atom reagents might be employed. There are reports in the literature of many attempts to make synthetic analogues of specific amino acids, by substituting selenium for sulfur residues in a chemically synthesized polypeptide or by removing an amino-terminal residue by the Edman technique and replacing it with an amino acid modified by a heavy atom [see Blundell & Johnson (1976) for a review]. Alternatively, analogues of the substrate of an enzyme or carrier protein can sometimes be modified with a heavy atom; however, this will disturb the active site, which is usually the region of greatest interest to the structural biologist. Such methods have not proved very useful and will not be described further here. Most proteins studied now are recombinant; site-directed mutagenesis can replace methionines in the sequence, which occur on average once every fifty residues, by selenomethionines (Hendrickson *et al.*, 1990) or more recently by telluromethionines (Budisa *et al.*, 1997). Such approaches have revolutionized macromolecular crystallography through the use of anomalous-dispersion techniques, but have yet to provide a very efficient method of introducing atoms heavier than selenium into proteins.

Thus, the vast majority of successful heavy-atom derivatives employed in crystallographic analyses are obtained on a trial-and-error basis. In earlier studies, the protein was often covalently modified, purified and characterized before crystallization. There are some useful covalent modifications, for example, the reaction of mercury with the sulfhydryl groups of cysteinyl side chains and the iodination of tyrosyl side chains. The replacement of a metal-ion cofactor, such as calcium or zinc, can also give a useful derivative. However, pre-reaction of the protein often gives rise to conformational changes in the protein, and crystallization frequently occurs in a different or non-isomorphous form.

Most heavy-atom derivatives are produced by direct soaking of the crystals in a solution of the heavy-atom compound. With this approach, heavy-atom substitution patterns tend to be complex, with sites frequently only partially occupied. The specificity is often determined by entropic factors. Thus, sites between molecules in the crystal lattice, or between several different side chains brought together by the tertiary structure, may bind the metal ion, even if the side chains individually do not have strong affinity for the metal. Chelation is entropically driven, and bonds may form with unusual protein ligands, a major factor causing lack of specificity.

Blake (1968) reviewed the data available for heavy-atom binding to proteins and suggested some generalizations. These were extended in a comprehensive review of protein heavy-atom derivatives (Blundell & Johnson, 1976; Blundell & Jenkins, 1977) which analysed the dependence of reactivity on protein side chain identity, nature of the reagent, pH, concentration, buffer *etc*. Over the past two decades, there have been discussions of the binding of some particular metal ions, but there have been no comprehensive analyses. Furthermore, protein–heavy-atom interactions have not often been fully described in publications of protein crystallographic analyses, and, in any case, the information has not been available in a format that could be used for systematic computer-based analysis.

We have now collected, either from the literature or directly from protein crystallographers, information on the preparation and characterization of heavy-atom derivatives of protein crystals. We have defined heavy atoms as those with an atomic weight greater than that of rubidium. We have assembled the information in the form of a data bank (Carvin *et al.*, 1991; Islam *et al.*, 1998) in which the coordinate data for the heavy-atom positions are compatible with the crystallographic data in the Protein Data Bank (Bernstein *et al.*, 1977). The data bank contains a wealth of information and provides the basis for further, more detailed analyses of heavy-atom binding to proteins. The information can be directly accessed and should be useful to protein crystallographers seeking to improve their success in preparing heavy-atom derivatives for isomorphous replacement and anomalous dispersion.

In this chapter we provide an introduction to the data bank and we review strategies that can be adopted in the preparation of heavy-atom derivatives of protein crystals for use in MIRAS.

12.1.2. Heavy-atom data bank

The heavy-atom data bank (HAD) is a computer-based archival file system that contains experimental and derived information from successful multiple isomorphous replacement analyses in the determination of protein crystal structures. HAD is available at http://www.bmm.icnet.uk/had/. The data bank makes available information which is otherwise only accessible in a widely distributed and fragmented form throughout the scientific literature or even unpublished in laboratory files.

The data bank contains information on heavy-atom derivatives for 969 protein crystals, 600 of which are deposited in the Protein Data Bank (PDB). A further 200 proteins are being processed at present. It contains information on the physical and chemical characteristics of each chemical compound that has proved successful in past protein crystallographic analyses: this includes the IUPAC name, trivial name, molecular formula, oxidation state, solution chemistry and stereochemistry. Experimental details of the preparation of the heavy-atom derivatives include the source of the protein, concentration of the heavy-atom solution, pH values, soak

times and details of the buffer compositions used in the experiments; 3164 different experimental conditions are recorded. The atomic coordinates are given in the same format as the PDB coordinates for the 5500 binding sites of the heavy atoms. A statistical analysis is included for each of the 456 heavy-atom reagents; this includes range of pH values and a summary of the amino acids involved at the binding sites. For metalloproteins, it gives details of the type, number, geometry of coordination and function of the native metal(s) present. This is followed by a description of the procedure for native-metal substitution and details of the coordination of the substituted heavy atom. It also includes an extensive bibliography and references to other relevant web sites.

12.1.3. Properties of heavy-atom compounds and their complexes

Potential ligands for heavy-atom reagents may be derived from the functional group(s) of reactive amino-acid side chains, from the buffer and from salting out/in agents. We must first consider factors that will influence the formation of such complexes in the environment of a protein crystal.

12.1.3.1. *Stability*

Ligands may be classified as either hard or soft. Hard ligands tend to be electronegative and interact electrostatically, with little delocalization of electron density. Water molecules, glutamates, aspartates, terminal carboxylates, and hydroxyl groups of serine and threonine from the protein, as well as acetate and citrate ions from the buffer, fall into this category. Conversely, soft ligands are polarizable and tend to form covalent bonds. Typical examples include the anions Cl^-, Br^-, I^-, S^{2-}, CN^-, imidazole, methionine, cysteine, cystine and histidine from the protein. Ligands can be listed in series of increasing hardness:

$$I^- > Br^- > Cl^- \gg F^- \gg H_2O,$$

$$RS^- > R_2S \gg NH_3 > H_2O,$$

$$CN^- > RNH_2 > Cl^- > CO_2^- > -\text{OH of alcohol}.$$

The metal components of the reagents may be classified as hard (class A) or soft (class B) in a similar way. Class A metals include the alkali metals, the alkaline earth metals, the lanthanide and actinide series, and the first-row transition metals from group III to group VA. Many of these metal ions have an inert-gas structure in which the electrons are held very strongly and tend to be non-polarizable. Metal ions in this class tend to interact with hard ligands, including the acetate, citrate and phosphate buffer components of mother liquor systems. On the other hand, class B metals have a preference for binding soft ligands. This group includes most members of the second and third row of the transition series (*e.g.* Ag, Cd, Pt, Au, Hg), which form cations such as $Pt(NH_3)_4^{2+}$ or anions such as $Au(CN)_2^-$, $PtCl_4^{2-}$ and HgI_4^{2-}. The easily polarizable *d* electrons allow formation of covalent bonds with methionine, cysteine and imidazole, so displacing the ligands of the complexes. In the middle and towards the end of the first transition-metal series, the ions have properties intermediate between class A and B metals. Class B character increases in the series:

$$Fe^{2+} < Co^{2+} < Ni^{2+} < Cu^{2+} < Zn^{2+}.$$

Thus, zinc binds to the polarizable sulfur of cysteine and imidazole of histidine as well as to carboxylates and water molecules. Tl^+ and Pb^{2+}, which each have an inert pair of electrons in their outer shell, are stable cations and prefer carboxylate rather than sulfur ligands or imidazole.

12.1.3.2. *Lability*

The rates at which ligands enter and leave a metal complex are important in the formation of heavy-atom derivatives, especially the covalent complexes of mercury, gold and platinum. The rate-determining step in unimolecular S_{N1} reactions is the expulsion of the leaving ligand from the metal complexes, which often proceeds relatively slowly. The intermediate complex, once formed, reacts with the entering ligand almost instantly. For S_{N1} reactions, the rate is directly proportional to the intermediate complex concentration but independent of the ligand concentration. The bimolecular S_{N2} mechanism involves attack by the ligand on the metal complex to form an intermediate complex, which then ejects the displaced ligand. The rate of reaction is proportional to the concentration of the initial species and the concentration of the nucleophile. S_{N2} reaction rates are dependent on the nature of the leaving group and the attacking nucleophile in the following ways:

$$\text{Relative rates of attack: } RS^- > I^- > Br^- > NH_3 > Cl^- > RO^-;$$

$$\text{Rate of leaving group: } H_2O > Cl^- > NO_2^- > CN^-.$$

Sulfur ligands are good nucleophiles but poor leaving groups. They form thermodynamically stable complexes. The rate of leaving is influenced by the *trans* effect in square-planar complexes of Au(III) and Pt(II). Thus groups in square-planar complexes *trans* to NH_3 are difficult to displace. This has implications for attempts to make derivatives of proteins in ammonium sulfate, where ligands may be replaced by NH_3.

Rates of reaction depend not only upon which ligands are present in a heavy-atom complex but also on the character of the metal. For example, $PtCl_4^{2-}$, $AuCl_4^-$ and $PdCl_4^-$ have similar square-planar geometries (Petsko *et al.*, 1978), but the rates of substitution vary:

$$PdCl_4^- > PtCl_4^{2-} > AuCl_4^-.$$

Thus, if the reaction between the protein and a palladium or platinum complex is proceeding too fast, a gold derivative might be investigated.

12.1.3.3. *Oxidation state of metal ions in protein crystals*

In the environment of a living cell, the following oxidation states tend to be stable:

$$\text{Os(II), Ru(II), Ir(III), Rh(III), Pt(II), Pd(II), Au(I).}$$

12.1.3.4. *Effect of pH*

Although the pK_a of an individual amino acid in solution is generally defined within narrow limits, environmental and steric factors give rise to a wide range of values in proteins. Thus, the hydrogen-ion concentration influences the thermodynamic and kinetic stability of potential complexes. Protons compete with heavy-atom ions for the available binding site(s) on the protein. For example, below pH 3.5, cations bind less well to aspartic and glutamic acids due to the protonation of the carboxylate groups.

The nucleophilicity of histidine increases when it loses its proton, and thus its positive charge changes from around pH 6.0 to 7.0. Similarly, the nucleophilicity of cysteine increases dramatically when the thiolate ion is formed at pH $\sim$ 8.0. The thiolate ion is a stronger nucleophile than the thioether group of methionine, but when it becomes protonated it is considerably less effective. The nucleophilicity of the attacking groups varies in the order

$$RS^- > R_2S > RSH.$$

references