International Tables for Crystallography (2006). Vol. F, Chapter 12.1, pp. 247–255.

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

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

Table 12.1.3.1. *Useful pH ranges of some heavy-atom reagents derived from the heavy-atom data bank*

No. of entries	Minimum	Average	Maximum	Compound
159	3.0	6.7	9.1	Potassium tetrachloroplatinum(II)
63	4.2	6.6	9.0	Potassium dicyanoaurate(I)
53	4.2	6.9	9.5	Mercury(II) chloride
59	2.8	6.7	9.0	Mercury(II) acetate
52	4.7	6.7	9.3	4-(Chloromercurio)benezenesulfonic acid
57	2.0	6.5	9.3	Potassium tetraiodomercurate(II)
36	5.4	6.7	8.5	Ethylmercurythiosalicylate (EMTS)
46	4.0	6.0	8.0	Potassium pentafluorooxyuranate(VI)
2	8.2	8.4	8.5	Barium(II) chloride
22	4.0	6.2	8.1	Lead(II) acetate
13	4.5	6.6	7.5	Lead(II) nitrate
1	6.5	6.5	6.5	Strontium(II) acetate
3	6.3	6.8	7.5	Thallium(I) acetate
2	5.9	6.6	7.2	Thallium(III) chloride
5	5.0	5.8	6.8	Gadolinium(III) chloride
9	4.9	6.7	7.5	Samarium(III) nitrate
7	4.9	6.6	8.7	Neodymium(III) chloride
64	4.1	6.3	8.6	Uranium(VI) oxyacetate

Thus the number and occupancy of sites can be manipulated by varying the pH, often after cross-linking the crystals to stabilize them.

Extremes in pH can give rise to considerable difficulties in establishing suitable derivatives, as hydrogen and hydroxyl ions compete with the metal ion/complex for the protein and with the protein for the metal ion/complex. At extremely high pH values metals in solution tend to form insoluble hydroxides. The ranges of pH values that are useful for metal ions are given in Table 12.1.3.1.

Varying the reactivity of amino-acid side chains by manipulation of the pH can enable the same heavy-atom ion/complex to bind at different sites, thus producing more than one derivative useful for phase determination.

12.1.3.5. *Effect of precipitants and buffers on heavy-atom binding*

Components present in the heavy-atom solution can have a profound effect on protein–heavy-atom interactions. The salting in/out agent (precipitant) and buffer are the principal sources of alternative ligands for the heavy-atom reagents, while protons compete with the heavy-atom ion/complex for the reactive amino-acid side chains.

Ammonium sulfate is the most successful precipitant in protein crystallization experiments (Gilliland *et al.*, 1994). However, its continued presence in the mother liquor can cause problems by interfering with protein–heavy-atom interactions. At high hydrogen-ion concentrations, the NH_3 group is protonated (*i.e.* NH_4^+), but as the pH rises the proton is lost, typically around pH 6.0–7.0, enabling the group to compete with the protein for the heavy-atom reagent by an S_{N2} reaction.

The nucleophilic strength of potential ligands follows the order

$$NH_3 > Cl^- > H_2O.$$

The anionic complex $PtCl_4^{2-}$ is present in excess ammonia at pH > 7.0 and it will react:

$$PtCl_4^{2-} \rightarrow \textit{cis}\text{-}PtCl_2(NH_3)_2 \rightarrow Pt(NH_3)_4^{2+}.$$

The resultant cationic complex is less susceptible to reaction due to the *trans* effect of NH_3. Pd, Au, Ag and Hg complexes react in a similar way. Decreasing the pH of the solution reduces the amount of free ammonia available through protonation (Sigler & Blow, 1965). Such a technique may give rise to other problems (*e.g.* cracked crystal, decreased nucleophilicity of the protein ligands).

Changing the precipitant to sodium/potassium phosphate or magnesium sulfate may alleviate the situation, but it may also present other problems. For instance, PO_4^{3-} displaces Cl^- from $PtCl_4^{2-}$, thus increasing the negative charge. Both PO_4^{3-} and SO_4^{2-} form insoluble complexes with class A metals (*e.g.* lanthanide and uranyl cations) (Petsko *et al.*, 1978). Both acetate and citrate form complexes with class A metals, but citrate, a chelating ion, binds more strongly. Tris buffer is probably preferable; it binds many cations, but the complexes formed tend to be relatively unstable.

12.1.3.6. *Solubility of heavy-atom compounds*

The solubility of a heavy-atom compound will depend upon the precipitant, buffer and pH. Typically, the component present in the highest concentration is the precipitant, either as salts (*e.g.* ammonium sulfate) or as an organic-based reagent (*e.g.* ethanol, MPD, PEG). Heavy-atom compounds that are essentially covalent and organic in character will be more soluble in ethanol, MPD, PEGs and other organic precipitants.

Although the solubility of tetrakis(acetoxymercurio)methane (TAMM) is higher than most multiple-heavy-atom compounds in aqueous solutions, the presence of glycylglycine or charged mercaptans, such as cysteamine or penicillamine, can increase solubility further (Lipka *et al.*, 1976). The ratio of TAMM to solubilization agent (*e.g.* glycylglycine) is typically 1:10. Even so, the final solubility of TAMM depends on the concentration of competing anions (*e.g.* chloride) (O'Halloran *et al.*, 1987).

Many organometallic compounds are relatively insoluble in aqueous solutions, but their solubility may be increased by pre-dissolving in an aprotic solvent such as acetonitrile.

Iodine and several inorganic iodide salts are insoluble in aqueous solutions. This can be rectified by dissolving the heavy-atom compounds in an aqueous solution of KI.

12.1.3.7. *Effect of concentration, time of soak and temperature on heavy-atom binding*

Most heavy-atom derivatives are prepared by diffusing or dialysing the compound into the crystals. Concentrations have typically ranged from 0.1–100.0 m*M*. Occasionally, concentrations as low as 0.001 m*M* have been employed to maintain crystal integrity. Low concentrations favour sites where the interactions between the heavy atom and the protein ligands are strongest. Decreasing the number of non-specific interactions minimizes the amount of heavy-atom reagent in the lattice. The latter absorbs X-rays without contributing to the diffraction pattern except at low angles. Increasing the concentration may give rise to other binding site(s). Usually, the higher the concentration employed, the shorter the soak time required for equivalent substitution. Short soak times at high concentrations tend to denature the crystals more often than long soaks at low concentrations. At very high concentrations (*i.e.* >100 m*M*), the heavy-atom compound perturbs the protein crystal–mother liquor equilibrium by withdrawing water molecules from the hydration shell around the periphery of the crystal. Disorder of the crystals can sometimes be avoided by the application of a cross-linking reagent (*e.g.* glutaraldehyde). The optimal concentration is the lowest concentration that consistently reproduces intensity differences in the diffraction pattern of 15–25% without cracking and disordering the crystals.

Length of soak may be important. The heavy-atom data bank shows that, typically, soak times range from one day to one week. Useful derivatives have been prepared with a soak time of an hour to over a year. If no binding is apparent after several days, extending the soak time to over a week may produce some binding, but this is rare. Soaks of 24 hours for simple inorganic salts and up to one week for other types of heavy-atom compounds will normally suffice when screening for binding. The concentration of the heavy-metal compound that can be achieved will depend on its solubility in the crystal stabilization solution. Normally, the longer the soak, the greater the occupancy. Exceptions can arise due to undesirable chemical reactions between components present in the derivatization solution.

For covalent-bond formation, the length of soak and the concentration can be short (*e.g.* 1 h, 0.01 m*M*). This is especially true for mercury derivatives of proteins that have reactive sulfhydryls (Ringe *et al.*, 1983).

Variations in the temperature can also alter the rate of reaction. The UO_2 acetate derivative of rhombohedral insulin binds twenty times more slowly at 4 °C than at ambient temperature (Blundell, 1968). A lower temperature allows greater control over the rate of substitution. Conversely, heavy-atom derivatives that do not appear to bind may do so upon elevation of the temperature.

12.1.4. Amino acids as ligands

The reactivity of the heavy-atom reagent will also depend on the state of the amino-acid residues in the protein.

The thiolate anion of *cysteine*, a potent nucleophile, reacts almost irreversibly with mercuric complexes or organomercurials. It also acts as a fast-entering attacking group in S_{N2} ligand substitution reactions with other class B metals (*e.g.* Ag, Ir, Rh, Pt, Pd, Au), forming stable complexes. Below pH 6, the thiolate anion becomes protonated. As covalent reactions are less sensitive to hydrogen-ion concentration than ligand substitution reactions, cysteines still bind rapidly with mercurials, but there is negligible reaction with other class B metals (Petsko *et al.*, 1978).

Cystines are very weakly reactive in ligand substitution reactions. However, $PtCl_4^{2-}$ binds to disulfides in some proteins with displacement of a chloride ion (Lipscomb *et al.*, 1970; Sigler *et al.*, 1968). Mercurials rarely insert spontaneously into disulfide linkages. However, substitution of mercury can be achieved either by the prior application of a reducing agent such as dithiothreitol (Ely *et al.*, 1973; Sperling *et al.*, 1969), or by direct application of reducing mercurous ions (Sperling & Steinberg, 1974).

The non-ionizable thioether group of *methionine* is unreactive towards mercurials, but the lone pair of electrons on sulfur allows nucleophilic S_{N2} ligand substitution. Methionine will displace Cl, I, Br and NO_2 ligands from platinum complexes to form a stable bond. The reaction of methionine with platinum compounds is not pH sensitive within the normal range. The residue may become unreactive through oxidation, first to the sulfoxide and then to the sulfone; only the sulfoxide can be reduced readily by thiols or other reducing agents.

Below pH 6, *histidine* exists mainly as an imidazolium cation. Although this is not reactive as a nucleophile, it can interact electrostatically with anionic complexes. At pH 7 and above, the unprotonated imidazole is a good nucleophile, being able to displace Cl, Br, I and NO_2 ligands from platinum, silver, mercury and gold complexes. Electrophilic substitution of iodine in the imidazole ring is feasible, but the conditions are severe and it has not proved very useful in preparing derivatives.

At pH < 8.5, the ε-amino group of *lysine* is protonated, allowing it to form weak electrostatic interactions with anionic heavy-atom complexes, but not to participate in S_{N2} substitution reactions. Above pH 9, the free amino group can displace Cl but not Br, I or NO_2 ligands from platinum and gold complexes. The pK_a of the guanidinium group of *arginine* is very high (>12 in proteins), so it interacts electrostatically as a cation with heavy-atom anionic complexes.

The indole ring of *tryptophan* is relatively inert to electrophilic substitution by iodine, but the ring nitrogen can be mercurated (Tsernoglou & Petsko, 1976). The reaction is not pH dependent, but there should be no competing nucleophiles in the mother liquor. Tryptophan does not usually participate as a ligand in substitution of heavy-atom complexes.

The phenolate oxygen anion of *tyrosine* is a good nucleophile and has the potential to bind a substantial number of heavy-atom complexes *via* S_{N2} ligand substitution reactions. However, it has a very high pK_a value of 10.5. Below pH 10, the protonated oxygen predominates, making electrophilic aromatic substitution by iodine the principal reaction.

Aspartic and glutamic acids have side-chain pK_a values in the range 3 to 4. At low pH, they will be protonated and unreactive. Above pH 5, the side chains will be anionic, making them good ligands for class A cations such as uranyl and rare earths. *Glutamine and asparagine* take part in metal coordination but rarely bind strongly enough to metal ligands on their own.

Hydroxyl groups of *serines and threonines* are fully protonated at normal pH values and are consequently not reactive nucleophiles. Abnormally reactive serines, usually at the active site as in serine proteases and β-lactamases, can react with heavy-atom reagents to give useful derivatives.

12.1.5. Protein chemistry of heavy-atom reagents

The heavy-atom data bank (Islam *et al.*, 1998) can be used to analyse the most commonly used heavy-atom reagents: these are given in Table 12.1.5.1. This shows that platinum, gold, mercury and uranyl have provided the most useful reagents.

The heavy-atom data bank can be used as a source of information about the reactivity of proteins to different heavy-atom reagents. This provides the basis for the following analysis.

Table 12.1.5.1. *The 23 most commonly used heavy-atom reagents*

The first column gives the number of times the reagent has been used in the analyses included in the heavy-atom data bank.

No.	Compound
287	Potassium tetrachloroplatinum(II)
111	Potassium dicyanoaurate(I)
103	Uranyl acetate
101	Mercury(II) acetate
98	Mercury(II) chloride
85	Ethylmercurythiosalicylate (EMTS)
82	Potassium tetraiodomercurate(II)
81	*para*-Chloromercuriobenzenesulfonate (PCMBS)
75	Trimethyllead(IV) acetate
73	Potassium pentafluorooxyuranate(VI)
73	Phosphatotris(ethylmercury)
61	Potassium tetranitritoplatinum(II)
60	Uranyl nitrate
58	Potassium tetracyanoplatinum(II)
57	Dichlorodiammineplatinum(II)
51	Potassium hexachloroplatinum(IV)
51	Methylmercury chloride
44	Potassium tetrachloroaurate(III)
42	*para*-Chloromercurybenzoate (PCMB)
39	Lead(II) acetate

12.1.5.1. *Hard cations*

Uranyl-ion complexes have proved the most popular A-group metal reagents for preparing heavy-atom derivatives of protein crystals (see Table 12.1.5.1). UO_2^{2+} is a linear, covalent group based on U(VI), the most stable oxidation state of uranium. Table 12.1.5.2 lists the most commonly used uranyl derivatives. Uranyl compounds may show $2+4$, $2+5$, or $2+6$ coordination, with ligands lying in or near a plane normal to the O═U═O^{2+} axis. These equatorial ligands may be neutral (*e.g.* H_2O) or anionic (*e.g.* NO_3^-, CH_3COO^-, oxalate^{2-}, F^-, Cl^- or O_2^-); the nitrate and acetate are bidentate ligands. An example is given in Fig. 12.1.5.1. Anionic complexes, such as $UO_2F_5^{3-}$, have been found near negatively charged amino-acid residues (*e.g.* Glu and Asp), suggesting that the equatorial ligands have been displaced. At low pH, uranyl groups have been located near the hydroxyl groups of threonine and serine residues.

The fifteen lanthanides have similar chemical properties and are generally used as nitrates, acetates or chlorides (Blundell & Johnson, 1976; Carvin, 1986). The lanthanide contraction, a steady decrease in size with increasing atomic number, allows selection of an ion with a radius that will give high occupancy and isomorphism. Gadolinium and samarium salts have the added advantage that the number of anomalous electrons is high.

Lanthanide ions have greater selectivity than the uranyl ion, which often forms clusters on the protein surface. Uranyl complexes and lanthanide ions are not very soluble above pH 7 and pH 9, respectively, due to the formation of hydroxides. Phosphate buffers should be avoided since they will compete for the heavy atom, often giving insoluble phosphates. In the presence of citrate, samarium is chelated and, since the citrate is difficult to replace, reaction may be inhibited. However, exchanging the buffer for Tris or acetate may enable a useful derivative to be obtained.

12.1.5.2. *Thallium and lead ions*

Thallium and lead can provide useful derivatives, especially in their lower oxidation states, Tl(I) and Pb(II), when they resemble class A metals. Owing to the non-group valence and presence of an inert pair of electrons, the ionic radii of Tl^+ (1.44 Å) and Pb^{2+} (1.21 Å) are greater than most class A metals. Thallous and plumbous cations prefer carboxylate rather than imidazole or sulfur ligands, although Pb^{2+} occasionally manifests its intermediate character by interacting with imidazole groups. Thallic (Tl^{3+}) and plumbic (Pb^{4+}) ions are similar to class B metals, showing preferential binding to soft ligands, but they are easily reduced in protein solutions.

12.1.5.3. *B-metal reagents*

The most useful members of the B-metal group, platinum, gold and mercury, give rise to an extensive range of heavy-atom compounds which form covalent, electrostatic and van der Waals complexes with proteins. Some compounds can bind to the protein molecule in different ways; for example, $PtCl_4^{2-}$ can bind either covalently to the thioether group of methionine or electrostatically with positively charged residues.

Mercury compounds have proved very successful for preparing heavy-atom derivatives of protein crystals (Table 12.1.5.1), mainly due to the ease of formation of covalent bonds with cysteine residues. An example is given in Fig. 12.1.5.2 in which mercuric chloride has been used to replace zinc in thermolysin. Hg^{2+} complexes are commonly two-coordinate linear and four-coordinate tetrahedral. The most popular mercury reagents are given in Table 12.1.5.3. The covalent character in Hg—*L* bonds, especially in the two-coordinate complexes, can cause solubility problems in aqueous solutions. However, an excess of an alkali metal salt (*e.g.* $HgI_2 + 2KI \rightarrow K_2HgI_4$) will often convert the compound to a more soluble anionic complex of the type HgX_4^{2-}, where $X = Cl^-$, Br^-, I^-, SCN^-, NCS^-, CN^-, SO_4^{2-}, oxalate^{2-}, NO_3^- or NO_2^-. In the presence of ammonium salts at high pH values, the cationic tetraammine complex, $Hg(NH_3)_4^{2+}$, tends to form. Variation in the

Table 12.1.5.2. *The five most popular uranium derivatives*

The first column gives the number of times the reagent has been used in the analyses included in the heavy-atom data bank.

No.	Compound
103	Uranyl acetate
73	Potassium pentafluorodioxyuranate(VI)
60	Uranyl nitrate
8	Uranium(VI) oxysulfate
4	Sodium triacetatedioxyuranate(VI)

Table 12.1.5.3. *The five most popular mercury derivatives*

The first column gives the number of times the reagent has been used in the analyses included in the heavy-atom data bank.

No.	Compound
101	Mercury(II) acetate
98	Mercury(II) chloride
85	Ethylmercurythiosalicylate (EMTS)
82	Potassium tetraiodomercurate(II)
81	*para*-Chloromercuriobenzenesulfonate (PCMBS)

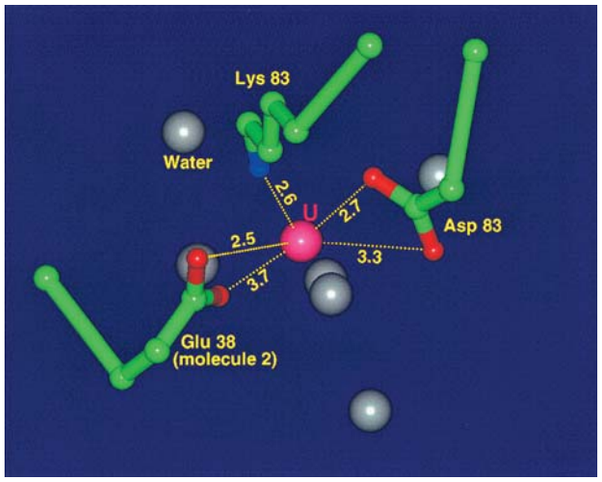

Fig. 12.1.5.1. The binding site for uranyl ions in cytochrome *b*5 (oxidized: 3B5C). The positions of the ligands in the parent crystals are shown; these probably move in the complex.

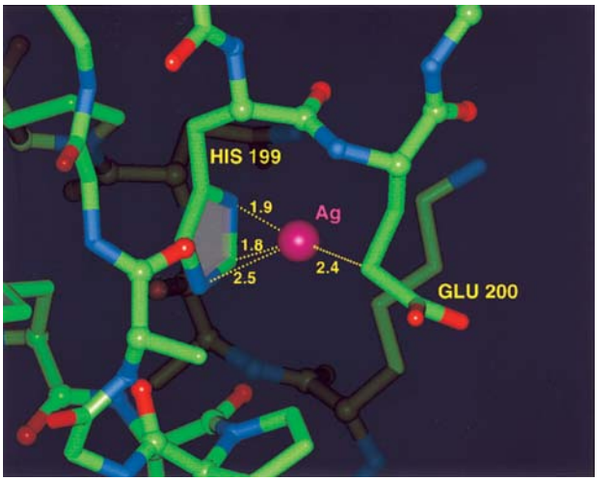

Fig. 12.1.5.3. The binding of a silver ion to immunoglobulin Fab (2FB4). The positions of the ligands in the parent crystals are shown, and these must move in the complex to coordinate the silver ion.

charge on the aromatic groups of organomercurials can give rise to different substitution patterns.

Silver, used as the nitrate, tends to interact with cysteine or histidine (see Fig. 12.1.5.3). In the presence of ammonium sulfate, it probably reacts as the ammonia complex, $Ag(NH_3)_4^+$. Silver ions are less polarizing and less reactive than Hg^{2+} ions; thus they give similar derivatives but often with less disorder, as in glucagon (Sasaki *et al.*, 1975). Where the metal ion displaces a proton, Ag^+ will need to react at a higher pH than Hg^{2+}.

The class B metals *palladium*, *platinum* and *gold* form stable covalent complexes with soft ligands, such as chloride, bromide, iodide, ammonia, imidazole and sulfur groups. The stereochemistry of their complexes depends on the number of *d* electrons present. For instance, the d^{10} ion of Au(I) gives a linear coordination of two [*e.g.* $Au(CN)_2^-$], whereas d^8 ions of Pd(II), Pt(II) and Au(III) are predominantly square planar, giving cationic [*e.g.* $Pt(NH_3)_4^{2+}$], anionic [*e.g.* $Au(CN)_4^-$, $PtCl_4^{2-}$ and $PdCl_4^{2-}$] or neutral [*e.g.* $Pt(NH_3)_2Cl_2$] complexes. These may accept an additional ligand to give square pyramidal coordination or two ligands to give octahedral coordination. The additional ligands are normally more weakly bound. Pt(IV) has a d^6 configuration and forms stable

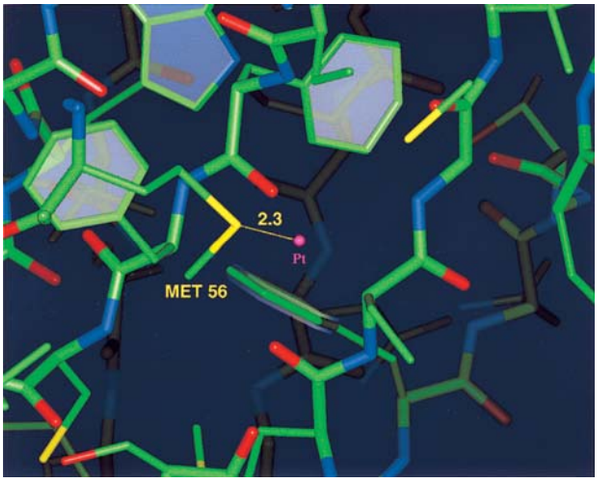

Fig. 12.1.5.4. The binding of $PtCl_4^{2-}$ through a methionine in azurin (1AZU).

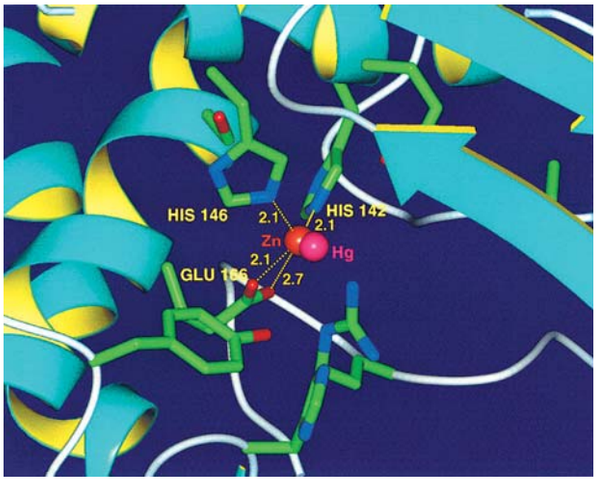

Fig. 12.1.5.2. Mercuric ions replace zinc in thermolysin (3TLN). The mercuric ion is shown superposed on the parent crystal structure; notice that the mercuric ion is slightly displaced from the zinc position due to its larger ionic radius.

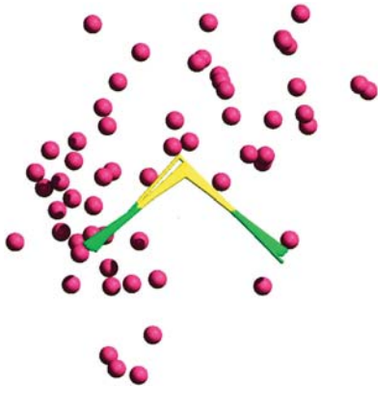

Fig. 12.1.5.5. The relative positions of methionine side chains (carbon: green; sulfur: yellow) in the parent crystals to the binding of platinum (pink) of $PtCl_4^{2-}$. The methionine side chains have been least-squares fitted.

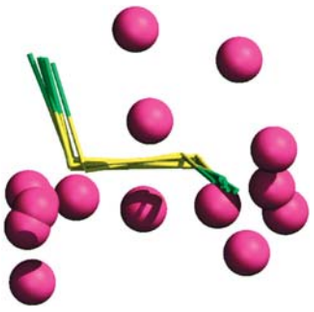

Fig. 12.1.5.6. The relative positions of cystine disulfide bridges (carbon: green; sulfur: yellow) in the parent crystals to the binding of platinum (pink) of $PtCl_4^{2-}$. The cystine side chains have been least-squares fitted, and only those with torsion angles in the range $99.7 \pm 8.3°$ have been used.

octahedral complexes, such as $PtCl_6^{2-}$, with six equivalent covalently bound ligands.

The kinetic and thermodynamic stability of these complexes depends on the protein ligands, buffer, pH and salting in/out agent (Petsko *et al.*, 1978). Anionic groups do not readily react with anionic reagents, such as RS^-, but are attacked more readily by neutral nucleophiles such as *R*SH, *R*-imidazole or RNH_2. The inert cationic group $Pt(NH_3)_4^{2+}$ is most likely to form electrostatic complexes with anionic groups, such as carboxylate. The neutral $Pt(NH_3)_2Cl_2$ molecule, however, can penetrate into hydrophobic areas but requires a stronger nucleophile such as RS^-. In acidic and neutral solutions, $PtCl_4^{2-}$ reacts most commonly with methionine (Figs. 12.1.5.4 and 12.1.5.5), cystine (disulfide) (Fig. 12.1.5.6), N-termini and histidine to form stable complexes. However, methionine reacts faster than histidine. Thus, it is possible to use time as a variable to define specificity. The most popular platinum reagents are listed in Table 12.1.5.4.

In aqueous solution, the square-planar complex $AuCl_4^-$ is hydrolysed to $Au(OH)_4^-$ in about one hour, or in the presence of a protein, reduced to Au(I) by methionine. In ammonium sulfate it probably exists as $AuCl_3(NH_3)$, $AuCl_2(NH_3)_2^+$ and $Au(NH_3)_4^{3+}$. In contrast, $Au(CN)_2^-$ is more stable and normally binds electrostatically. However, on occasions at pH > 6.0, the $Au(CN)_2^-$ complex has bound to cysteine residues by nucleophilic displacement reactions.

Osmium resembles platinum in many ways and typically acts as a class B metal. It occurs in all oxidation states from 0 to VIII, but most usually in III, as in K_3OsCl_6; in IV, as in K_2OsCl_6; in VI, as in $K_2OsO_2(OH)_4$; and in VIII, as in osmium tetraoxide, OsO_4. Higher-oxidation-state compounds tend to be reduced to $OsO_2(OH)_2$ in most crystallization solutions and in the presence of ammonia or halide ion they can become further reduced to cationic or anionic complexes, such as $Os(NH_3)_6^{3+}$ or $OsCl_6^{2-}$. Anionic complexes may be substituted by histidine residues at pH > 7.0 or bound as ion pairs by histidine at pH < 7.0 or protonated amino groups. Cationic complexes tend to bind to negatively charged residues *via* electrostatic interactions.

Iridium is found in all oxidation states from II to VI but commonly exists in III, as in K_3IrCl_6, and IV, as in $(NH_4)_2IrCl_6$. Ir(III) is similar to *rhodium*(III) and is found in a variety of cationic, uncharged and anionic complexes. All Ir(III) complexes are kinetically inert, whereas most anionic complexes of Rh(III) are labile. Ir(IV) is commonly found as the hexahalo complexes IrX_6^{2-} (except iodine), which are also fairly kinetically inert. Cationic [*e.g.* $Ir(NH_3)_6^{3+}$], neutral (*i.e.* $IrCl_3$) and anionic (*i.e.* $IrCl_6^{2-}$) species have proved useful in forming derivatives of protein crystals.

12.1.5.4. *Electrostatic binding of heavy-atom anions*

Positively charged groups of proteins, such as the α-amino terminus, ε-amino of lysine, guanidinium of arginine and imadazolium of histidine, may form ion pairs with heavy-atom anionic complexes. For example, HgI_4^{2-} and HgI_3^- can bind through electrostatic interactions. Anionic metal cyanide complexes tend to be more resistant to substitution and consequently interact electrostatically on most occasions. For example, $Pt(CN)_4^{2-}$ binds at several sites involving lysine or arginine residues in proteins (Fig. 12.1.5.7). $Pt(CN)_4^{2-}$ and $Au(CN)_2^-$ can act as inhibitors by binding at coenzyme phosphate sites.

12.1.5.5. *Hydrophobic heavy-atom reagents*

Since many heavy-atom reagents are hydrophilic, most interactions occur at the protein surface. However, substitution, addition or removal of the non-heavy-atom component(s) of the reagent can alter the hydrophilic–hydrophobic balance and lead to penetration of the core. For example, anionic complexes such as $HgCl_4^{2-}$ and $PbCl_6^{2-}$ are hydrophilic and would not normally enter the protein core, although organometallics, such as *R*HgCl and R_3PbCl (*R* = aliphatic or aromatic), are much more hydrophobic and can do so.

Hydrophobic organomercury compounds of the general formula *R*Hg*X*, where *R* is an aliphatic or aromatic organic group, react with sulfhydryls through displacement of *X*. When *X* is PO_4^{3-}, SO_4^{2-} or NO_3^-, the bond is ionic, making the formation of the cation RHg^+ easier. *R* is often chosen to be a small aliphatic group (*e.g.* CH_3, C_2H_5). However, the presence of a benzene ring enhances the

Table 12.1.5.4. *The five most popular platinum derivatives*

The first column gives the number of times the reagent has been used the analyses included in the heavy-atom data bank.

No.	Compound
287	Potassium tetrachloroplatinum(II)
61	Potassium tetranitritoplatinum(II)
58	Potassium tetracyanoplatinum(II)
57	Dichlorodiammineplatinum(II)
51	Potassium hexachloroplatinum(IV)

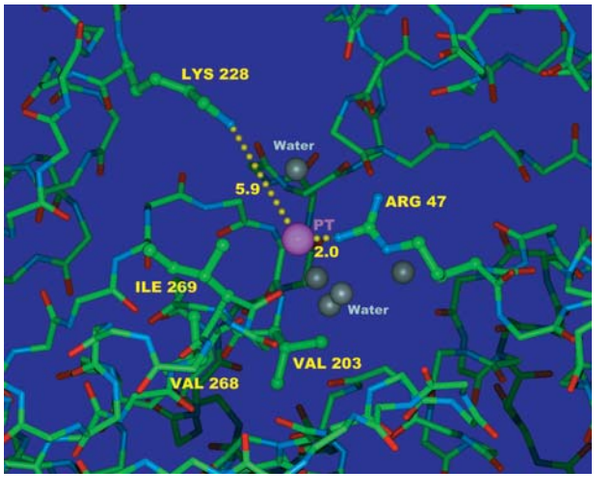

Fig. 12.1.5.7. The binding of $Pt(CN)_4^{2-}$ to aldose dehydrogenase (8ADH).

stability of the heavy-atom reagent. Careful selection of the *X* group can assist penetration into the hydrophobic core. The hydrophobicity of *X* follows the order

$$PO_4^{3-} < NO_3^- < Cl < Br < I < R.$$

*R*Hg*R* (*R* = aliphatic or aromatic) compounds also bind sulfhydryl residues in hydrophobic regions. The mechanism of reaction of methylphenylmercury with buried sulfhydryl groups may involve fast dissolution in the hydrophobic interior of the protein followed by a slow reaction with neighbouring sulfhydryl residues (Abraham *et al.*, 1983). They are difficult to prepare in aqueous solutions; an aprotic solvent, such as acetonitrile, can improve solubility, but this is not normally a problem in high concentrations of organic components, such as PEG, MPD or ethanol.

Inert gases were first used in the analysis of myoglobin. Schoenborn *et al.* (1965) discovered that the hydrophobic site that bound HgI_3^- also bound a xenon atom at 2.5 atmospheres. They proposed that this may be a general way of producing heavy-atom derivatives of proteins. Recently, there has been increasing interest in this idea, which has now been developed to produce well defined derivatives of a wide range of different proteins. Crystals are subjected to high gas pressures. Xenon requires about 10 atmospheres in order to get saturated binding sites. Krypton binds much less strongly and requires around 60 atmospheres. Since the binding of both inert gases is reversible, it is necessary to keep the protein crystals in a gaseous environment in a specialized pressure cell. Such pressure cells have been developed by Schiltz (1997) at LURE. Xenon binds to hydrophobic cavities, with little conformational change and a retention of isomorphism in crystals. Krypton binds at the same sites as xenon, but since it is lighter and needs higher pressure it has been exploited less by protein crystallographers. However, it has a well defined *K* edge at around 1 Å and so has attractions for multiple-wavelength anomalous dispersion.

12.1.5.6. *Iodine*

In addition to their use in isomorphous replacement, iodine derivatives of crystalline proteins have been prepared as tyrosine or histidine markers to assist main-chain tracing and to act as a probe for surface residues. The order of reactivity towards these reactive residues is

$$\text{Tyr} > \text{His} \gg \text{Trp}.$$

I_3^-, I^-, I^+ and I_2 can be generated by several different methods. An equimolar solution of KI/I_2 or NaI/I_2 in 5% (*v*/*v*) ethanol/water solution is often used to generate the anionic species I_3^- and I^-. An oxidizing agent, such as chloramine T, can be added to KI, typically in a concentration ratio of 1:50; alternatively, polystyrene beads derivatized with *N*-chlorobenzene sulfonamide can be used with NaI. Similarly, the addition of excess KI to ICl or OI^- will generate I_3^-, I^- and I^+. To avoid oxidation of iodine solutions, the pH should be less than 5.0. To avoid cracking the crystals, it may be necessary to increase the iodine concentration very slowly and to wash the derivatized crystals in the mother liquor in order to remove free I_2. Mono- or di-iodination of tyrosines can cause disruption of the protein structure either because of the larger size or the breaking of hydrogen bonds due to lowering of the pK_a of the phenolic hydroxyl.

12.1.5.7. *Polynuclear reagents*

The structure determination of large multicomponent systems such as the 50S ribosomal subunit (Yonath *et al.*, 1986) or the nucleosome core particle (O'Halloran *et al.*, 1987) requires the addition of reagents with a greater number of electrons, preferably in a compact polynuclear structure. Such reagents may be either cluster compounds or multimetal centres having metal–metal bonds.

Polynuclear reagents should preferably be covalently bound to one or a few specific sites, either first in solution or later in the crystals. Spacers of differing length can be inserted into the reagent to increase accessibility. Their low solubility in aqueous solutions can often be overcome by dissolving them in an apolar solvent (*e.g.* acetonitrile). Tetrakis(acetoxymercurio)methane (TAMM) and di-*m*-iodobis(ethylenediamine)diplatinum(II) nitrate (PIP) have better solubility in aqueous solutions than other polynuclear heavy-atom compounds.

Polynuclear heavy-atom reagents give an enhanced signal-to-noise ratio in low-resolution MIR studies, but this advantage is offset by the fall-off in scattering amplitude that arises from interference of diffracted waves at higher resolution. In the nucleosome core particle, the scattering reached 50% of its zero-angle value at 7.0 Å, while the relative drop for a single heavy atom was 10% (O'Halloran *et al.*, 1987). Cluster and multimetal reagents that have been successfully employed in protein structure determinations have been reviewed by Thygesen *et al.* (1996).

12.1.6. Metal-ion replacement in metalloproteins

The metal-ion cofactor can sometimes be displaced by dialysis or diffusion by a heavy-atom solution, but usually the cofactor is first removed by a chelating agent (*e.g.* EDTA) or by acidification. These are best carried out on the crystals. Alternatively, the metal

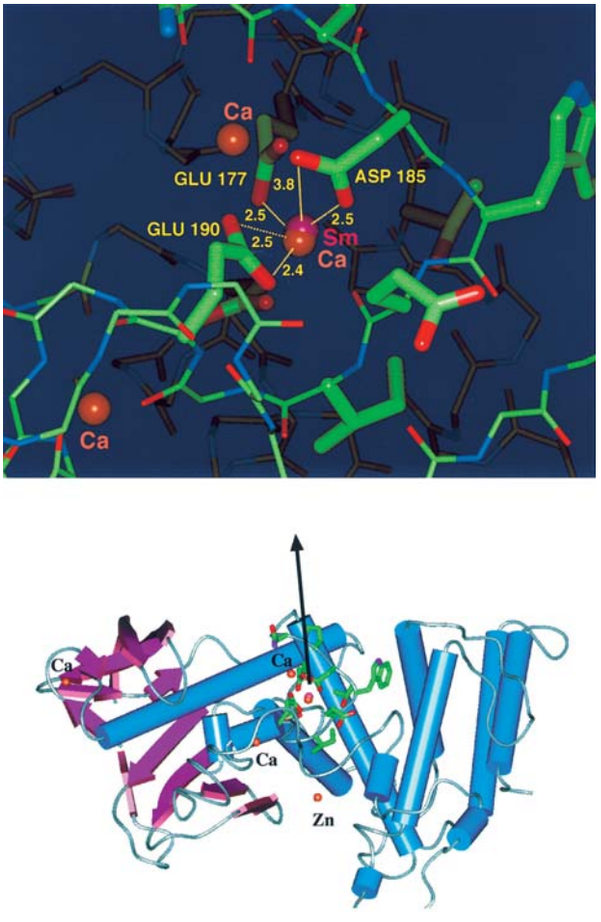

Fig. 12.1.6.1. The displacement of calcium by samarium in thermolysin. The samarium of the heavy-atom derivative is shown superposed on the parent crystal structure.

can be substituted by biosynthesis of the metalloprotein under enriched conditions of the substituting metal, an approach which has been successful in displacing zinc with cobalt and other lighter metals.

The metal ions are best substituted by a metal of similar character and radius. Thus, calcium is an A-group metal which prefers ligands containing oxygen atoms that may originate from carboxylic, carboxyamide, hydroxyl, main-chain carbonyl groups and water molecules. Divalent alkaline earth metal ions (*e.g.* Sr^{2+}, Ba^{2+}) or trivalent lanthanide ions can bind at calcium sites but can give very different coordination geometry and stability. Nd^{3+} and Sm^{3+} can displace some Ca^{2+} ions with negligible change in structure (Fig. 12.1.6.1). On the other hand, zinc has a relatively small ionic radius and is more polarizing. Structural zinc atoms are often tetrahedrally coordinated by cysteine residues, while those at active sites frequently bind histidine, often in association with a water molecule and/or carboxylate ligands. Cadmium or mercury can replace zinc, but often with a conformational change leading to lack of isomorphism.

12.1.7. Analogues of amino acids

Attempts to replace amino acids by heavy-atom substituted synthetic analogues with a similar charge and shape have not proved successful in large proteins, although a selenocystine was used successfully in the analysis of oxytocin (Wood *et al.*, 1986). However, the production of proteins labelled by selenium using biological substitution of selenomethionine (SeMet) for methionine (Hendrickson, 1985) has been stimulated by multiple-wavelength anomalous dispersion (MAD) (Hendrickson *et al.*, 1990). Methionine biosynthesis is blocked in the cells in which the protein is produced and SeMet is substituted for Met in the growth medium. The generality of the labelling scheme for proteins is the root of its success, as discussed by Doublié (1997).

SeMet has been incorporated into proteins expressed in *Escherichia coli* strains that are auxotrophic for Met [strain DL421 (Hendrickson *et al.*, 1990); strain B834 (Leahy *et al.*, 1994); strain LE392 (Ceska *et al.*, 1996)]. Nearly complete incorporation has also been reported in non-auxotrophic bacterial strains, *E. coli* strain XA90 (Labahn *et al.*, 1996), in a mammalian cell line (Lustbader *et al.*, 1995) and in baculovirus-infected insect cells (Chen & Bahl, 1991). Usually, somewhat higher than normal concentrations of disulfide reducing agents, such as dithiothreitol or mercaptoethanol, are sufficient to protect SeMet from air oxidation to the selenoxide, although crystallization in an inert atmosphere may be necessary. Proteins usually have SeMet substituted for Met at levels approaching 100%. The cells are viable and the proteins are functional.

Site-directed mutagenesis offers an alternative approach for the introduction of specific heavy-atom binding sites. A common procedure is to replace residue(s) in the variable part of the primary structure with cysteine. The selection of the residue to mutate in a protein of unknown structure remains a challenge.

Although selenocysteine is toxic to cells, cysteine auxotrophic strains, in which proteins can be synthesized with the seleno derivative, have been developed (Miller, 1972; Muller *et al.*, 1994). The bacteria are grown under limiting amounts of cysteine with no other sulfur source. They are induced for 10 min and then resuspended in selenocysteine for a 3 h incubation. The protein is purified with a reducing agent. In general, the substitution at the selenocysteine seems to be less satisfactory than selenomethionine, with occupancy often as low as 20%.

Budisa *et al.* (1997) have experimented on incorporating a range of novel amino-acid analogues using *in vitro* suppression. This is achieved by suppressing the stop colons and engineering tRNA synthases to incorporate the analogue. Possible candidates are telluromethionine, 5-bromotryptophan, 5-iodotryptophan, selenotryptophan and tellurotryptophan. The bioincorporation of TeMet into derivatized crystals did not greatly affect their stability in buffer solutions and to X-radiation. Isomorphism was maintained despite the C—Te bond being longer than C—Se or C—S. TeMet crystals are not as suitable for MAD analysis as SeMet crystals due to the 0.3 Å absorption edge of tellurium. The method is restricted to methionine residues located in the hydrophobic regions, since solvent accessibility may cause undefined chemical reactions with the highly reactive C—Te side chain. Thus the protein must be expressed in the folded form.

12.1.8. Use of the heavy-atom data bank to select derivatives

The heavy-atom data bank is probably best exploited by first investigating the most commonly used heavy-atom reagents with a view to obtaining mercury, platinum and uranyl derivatives that tend to bind at different sites. The most common reagents (Table 12.1.5.1) can first be selected and tested for suitability in terms of amino-acid sequence, pH, buffer and salt. If there are many sulfhydryls, several mercurials might be exploited, or if there are several methionines, other platinum agents might be investigated. A high pH would argue against use of uranyl due to the insolubility of hydroxides; the presence of ammonium sulfate would argue for as low a pH as possible. The presence of citrate would imply changing the buffer for acetate if A-group metals, such as uranium or lanthanides, were to be used.

For each heavy-atom agent, the conditions of its previous use can be checked against the conditions of crystallization in the current study. Conversely, the database can be interrogated for reagents that have been used in similar conditions. In each case, derivatives that maximize the variety of ligands should be exploited.

The time of soak should be first set according to previous experience indicated in the database. However, the progress of heavy-atom substitution needs to be monitored by checking for change of colour, transparency or cracking. If cracking and disruption of the crystals occurs quickly, a less reactive reagent can be tried, and, conversely if substitution is insufficient, a more reactive reagent can be tried. If there are several cysteines, different derivatives can be obtained with mercurials of different size and hydrophobicity. In each circumstance, the data bank should provide useful information to assist in choosing reagents.

Please keep information about the heavy-atom binding sites and the heavy-atom structure-factor amplitudes. They should be submitted to the Protein Data Bank.

Acknowledgements

We thank all those who have generously sought out and sent us details of the heavy-atom binding sites in their derivatives, and the ICRF and the Wellcome Trust for support.

crystal form, which in turn provided phases of sufficient quality to determine heavy-atom sites in derivatives of the second form. Phase-combination and density-modification techniques in the two crystal forms allowed the solution of the structure.

12.2.5.2. *Space-group problems*

Although the macromolecular crystallographer is rarely confronted with the problems facing their small-molecule colleagues with regard to determining the correct space group, the simplified heavy-atom structure may often throw some surprises. Certain pseudosymmetries may become 'exact' for the heavy-atom difference Patterson map. Thus, cross peaks between different heavy atoms may occur on a Harker section (or 'pseudo-Harker section'), complicating interpretation of the Patterson map. Such was the case with azurin (Adman *et al.*, 1978; Nar *et al.*, 1991), where the heavy-atom structure gave rise to a pseudo-homometric Patterson function, *i.e.* one in which two possible (nonequivalent) choices were available for the heavy-atom structure, only one of which was correct. This arose from a pseudo-centring of the lattice that became almost exact for the heavy-atom structure.

In the case of human NC1 (Stubbs *et al.*, 1990), all heavy-atom derivatives appeared to lie on or near the crystallographic twofold axis. This resulted in a partially centrosymmetric heavy-atom structure that failed to deliver sufficient phase information for noncentrosymmetric reflections. To check for problems with the native data set, anomalous difference Patterson maps $\{\text{coefficients } [F_{PH}(\mathbf{S}) - F_{PH}(-\mathbf{S})]^2\}$ were calculated. Coincidence of the peaks obtained from conventional and anomalous Patterson syntheses showed that the heavy-atom positions were correct, but unfortunately did not lead to a structure solution.

12.2.5.3. *High levels of substitution; noncrystallographic symmetry*

Most problematic are the cases where many heavy atoms have become incorporated in the asymmetric unit. Not only does this cause difficulties in the scaling of derivative to native data, but also the large number of peaks results in ambiguities in the solution of the Patterson function. In such cases, it may be necessary to obtain primary phase information from a different source (such as, for example, another low-substitution-site derivative). One important subclass of high-level substitution is when the native asymmetric unit contains several copies of a single molecule (noncrystallographic symmetry or NCS).

A major problem in locating complex noncrystallographic axes is that the geometrical relationship between NCS peaks in the Patterson map is nontrivial. Under certain conditions, NCS results in a recognizable local symmetry within the Patterson map (Stubbs *et al.*, 1996). In many cases, however, these conditions (that the NCS axes of crystallographic symmetry-related molecules are parallel) are not fulfilled. Under such circumstances, all heavy-atom sites (including all crystallographic symmetry-related positions) must be checked carefully with the rotation function in order to pinpoint the NCS axis. This is relatively trivial for low-order NCS (twofold, threefold), but becomes increasingly complicated for higher orders. It should also always be borne in mind that the heavy-atom positions might not necessarily follow the NCS constraints due to crystal packing. If there is reason to suspect that sites are related by local symmetry, then the orientation of this axis can be used in the initial Harker searches; in practice, however, such searches are extremely sensitive to the correct orientation of the axis.

In the case of high-order NCS (such as, *e.g.*, with icosahedral virus structures or symmetric macromolecular complexes), an alternative approach to the usual initial Harker-vector search can be provided by the self-rotation function. Knowledge of the orientation of the NCS axis (from the rotation function) can be used to determine the relative positions of heavy atoms to the NCS axis (Argos & Rossmann, 1976; Arnold *et al.*, 1987; Tong & Rossmann, 1993). The orientation can be refined and the resulting peaks can be used as input in a subsequent translation search of the Harker sections.

References

12.1

Abraham, D. J., Phillips, S. E. V. & Kennedy, P. E. (1983). *Methylphenylmercury: a novel heavy atom reagent for protein crystallography. J. Mol. Biol.* **170**, 249–252.

Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F. Jr, Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). *The Protein Data Bank: a computer-based archival file for macromolecular structures. J. Mol. Biol.* **112**, 534–552.

Blake, C. C. F. (1968). *The preparation of isomorphous derivatives.* In *Advances in protein chemistry,* Vol. 23, pp. 59–120. New York and London: Academic Press.

Blundell, T. L. (1968). Unpublished results.

Blundell, T. L. & Jenkins, J. A. (1977). *The binding of heavy metals to proteins. Chem. Soc. Rev. (London)*, **6**, 139–171.

Blundell, T. L. & Johnson, L. N. (1976). *Protein crystallography.* New York: Academic Press.

Budisa, N., Karnbrock, W., Steinbacher, S., Humm, A., Prade, L., Neuefeind, T., Moroder, L. & Huber, R. (1997). *Bioincorporation of telluromethionine into proteins: a promising new approach for X-ray structure analysis of proteins. J. Mol. Biol.* **271**, 1–8.

Carvin, D. G. A. (1986). Unpublished results.

Carvin, D. G. A., Islam, S. A., Sternberg, M. J. E. & Blundell, T. L. (1991). *Isomorphous replacement and anomalous scattering*. Warrington: Daresbury Laboratory.

Ceska, T. A., Sayers, J. R., Stier, G. & Suck, D. (1996). *A helical arch allowing single-stranded DNA to thread through t5 5′-exonuclease. Nature (London)*, **382**, 90–93.

Chen, W. & Bahl, O. P. (1991). *Recombinant carbohydrate and selenomethionyl variants of human choriogonadotropin. J. Biol. Chem.* **266**, 8192–8197.

Doublié, S. (1997). *Preparation of selenomethionyl proteins for phase determination. Methods Enzymol.* **276**, 523–530.

Ely, K. R., Girling, R. L., Schiffer, M., Cunningham, D. E. & Edmundson, A. B. (1973). *Preparation and properties of crystals of a Bence–Jones dimer with mercury inserted into the interchain disulphide bond. Biochemistry*, **12**, 4233–4237.

Gilliland, G. L., Tung, M., Blakeslee, D. M. & Ladner, J. E. (1994). *Biological Macromolecular Crystallization Database, version 3.0: new features, data and the NASA archive for protein crystal growth data. Acta Cryst.* D**50**, 408–413.

Green, D. W., Ingram, V. M. & Perutz, M. F. (1954). *The method of isomorphous replacement for protein crystallography. Proc. R. Soc. London Ser. A*, **225**, 287–307.

Hendrickson, W. A. (1985). *Analysis of protein structure from diffraction measurement at multiple wavelengths. Trans. Am. Crystallogr. Assoc.* **21**, 11–21.

12.1 (*cont.*)

Hendrickson, W. A., Horton, J. R. & LeMaster, D. M. (1990). *Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. EMBO J.* **9**, 1665–1672.

Islam, S. A., Carvin, D., Sternberg, M. J. E. & Blundell, T. L. (1998). *HAD, a data bank of heavy-atom binding sites in protein crystals: a resource for use in multiple isomorphous replacement and anomalous scattering. Acta Cryst.* D**54**, 1199–1206.

Labahn, J., Scharer, O. D., Long, A., Ezaz-Nikpay, K., Verdine, O. L. & Ellenberger, T. E. (1996). *Structural basis for the excision repair of alkylation-damaged DNA. Cell*, **86**, 321–329.

Leahy, D. J., Erickson, H. P., Aukhil, I., Joshi, P. & Hendrickson, W. A. (1994). *Crystallization of a fragment of human fibronectin: introduction of methionine by site-directed mutagenesis to allow phasing via selenomethionine. Proteins Struct. Funct. Genet.* **19**, 48–54.

Lipka, J. J., Lippard, S. J. & Wall, J. S. (1976). *Visualisation of polymercurimethane-labelled fd bacteriophage in the scanning transmission electron microscope. Science*, **206**, 1419–1421.

Lipscomb, W. N., Reeke, G. N., Hartsuck, J. A., Quiocho, F. A. & Bethge, P. H. (1970). *The structure of carboxypeptidase a. VIII. Atomic interpretation at 0.2 nm resolution, a new study of the complex of glycyl-L-tyrosine with CPA, and mechanistic deductions. Philos. Trans. R. Soc. London Ser. B*, **257**, 177–214.

Lustbader, J. W., Wu, H., Birken, S., Pollak, S., Kolks-Gawinowicz, M. A., Pound, A. M., Austen, D., Hendrickson, W. A. & Canfield, R. E. (1995). *The expression, characterization and crystallization of wild-type and selenomethionyl human chorionic gonadotrophin. Endocrinology*, **136**, 640–650.

Miller, J. H. (1972). *Cysteine auxotrophic strains, in which proteins can be synthesised with the seleno derivative.* In *Experiments in molecular genetics.* Cold Spring Harbour Laboratory Press.

Muller, S., Senn, H., Gsell, B., Vetter, W., Baron, C. & Bock, A. (1994). *The formation of diselenide bridges in proteins by incorporation of selenocysteine residues: biosynthesis and characterization of (Se)2-thioredoxin. Biochemistry*, **33**, 3404–3412.

O'Halloran, T. V., Lippard, S. J., Richmond, T. J. & Klug, A. (1987). *Multiple heavy-atom reagents for macromolecular X-ray structure determination application to the nucleosome core particle. J. Mol. Biol.* **194**, 705–712.

Petsko, G. A., Phillips, D. C., Williams, R. J. P. & Wilson, I. A. (1978). *On the protein crystal chemistry of chloroplatinite ions: general principles and interactions with triose phosphate isomerase. J. Mol. Biol.* **120**, 345–359.

Ringe, D., Petsko, G. A., Yamakura, F., Suzuki, K. & Ohmori, D. (1983). *Structure of iron superoxide dismutase from Pseudomonas ovalis at 2.9 Å resolution. Proc. Natl Acad. Sci. USA*, **80**, 3879–3883.

Sasaki, K., Dockerill, S., Adamiak, D. A., Tickle, I. J. & Blundell, T. L. (1975). *X-ray analysis of glucagon and its relationship to receptor binding. Nature (London)*, **257**, 751–757.

Schiltz, M. (1997). *Xenon & krypton at LURE.* http://www.lure.u-psud.fr/sections/Xenon/XENONENG.HTM.

Schoenborn, B. P., Watson, H. C. & Kendrew, J. C. (1965). *Binding of xenon to sperm whale myoglobin. Nature (London)*, **207**, 28–30.

Sigler, P. B. & Blow, D. M. (1965). *A means of promoting heavy atom binding in protein crystals. J. Mol. Biol.* **14**, 640–644.

Sigler, P. B., Blow, D. M., Matthews, B. W. & Henderson, R. (1968). *Structure of crystalline alpha-chymotrypsin II. A preliminary report including a hypothesis for the activation mechanism. J. Mol. Biol.* **35**, 143–164.

Sperling, R., Burstein, Y. & Steinberg, I. Z. (1969). *Selective reduction and mercuration of cysteine IV–V in bovine pancreatic ribonuclease. Biochemistry*, **8**, 3810–3820.

Sperling, R. & Steinberg, I. Z. (1974). *Simultaneous reduction and mercuration of disulphide bond A6–A11 of insulin by monovalent mercury. Biochemistry*, **13**, 2007–2013.

Thygesen, J., Weinstein, S., Franceschi, F. & Yonath, A. (1996). *The suitability of metal clusters for phasing in macromolecular crystallography of large macromolecular assemblies. Structure*, **4**, 513–518.

Tsernoglou, D. & Petsko, G.-A. (1976). *The crystal structure of a post-synaptic neurotoxin from sea snake at 2.2 Å resolution. FEBS Lett.* **68**, 1–4.

Wood, S. P., Tickle, I. J., Treharne, A. M., Pitts, J. E., Mascarenhas, Y., Li, J. Y., Husain, J., Cooper, S., Blundell, T. L., Hruby, V. J., Buku, A., Fischman, A. J. & Wyssbrod, H. R. (1986). *Crystal structure analysis of deamino-oxytocin: conformational flexibility and receptor binding. Science*, **232**, 633–636.

Yonath, A., Saper, M. A., Makowski, I., Mussig, J., Piefke, J., Bartunik, H. D., Bartels, K. S. & Wittmann, H. G. (1986). *Characterization of single crystals of the large ribosomal particles from bacillus stearothermophilus. J. Mol. Biol.* **187**, 633–636.

12.2

Adman, E. T., Stenkamp, R. E., Sieker, L. C. & Jensen, L. H. (1978). *A crystallographic model for azurin at 3.0 Å resolution. J. Mol. Biol.* **123**, 35–47.

Argos, P. & Rossmann, M. G. (1976). *A method to determine heavy-atom positions for virus structures. Acta Cryst.* B**32**, 2975–2983.

Arnold, E., Vriend, G., Luo, M., Griffith, J. P., Kamer, G., Erickson, J. W., Johnson, J. E. & Rossmann, M. G. (1987). *The structure determination of a common cold virus, human rhinovirus 14. Acta Cryst.* A**43**, 346–361.

Badger, J. & Athay, R. (1998). *Automated and graphical methods for locating heavy-atom sites for isomorphous replacement and multiwavelength anomalous diffraction phase determination. J. Appl. Cryst.* **31**, 270–274.

Blow, D. M. & Crick, F. H. C. (1959). *The treatment of errors in the isomorphous replacement method. Acta Cryst.* **12**, 794–802.

Budisa, N., Karnbrock, W., Steinbacher, S., Humm, A., Prade, L., Neuefeind, T., Moroder, L. & Huber, R. (1997). *Bioincorporation of telluromethionine into proteins: a promising new approach for X-ray structure analysis of proteins. J. Mol. Biol.* **270**, 616–623.

Buerger, M. J. (1959). *Vector space*. New York: Wiley.

Chang, G. & Lewis, M. (1994). *Using genetic algorithms for solving heavy-atom sites. Acta Cryst.* D**50**, 667–674.

Crick, F. H. C. & Magdoff, B. S. (1956). *The theory of the method of isomorphous replacement for protein crystals. I. Acta Cryst.* **9**, 901–908.

Dumas, P. (1994*a*). *The heavy-atom problem: a statistical analysis. I. A priori determination of best scaling, level of substitution, lack of isomorphism and phasing power. Acta Cryst.* A**50**, 526–537.

Dumas, P. (1994*b*). *The heavy-atom problem: a statistical analysis. II. Consequences of the a priori knowledge of the noise and heavy-atom powers and use of a correlation function for heavy-atom-site determination. Acta Cryst.* A**50**, 537–546; erratum, A**50**, 793.

Green, D. W., Ingram, V. M. & Perutz, M. F. (1954). *The structure of haemoglobin IV. Sign determination by the isomorphous replacement method. Proc. R. Soc. London Ser. A*, **225**, 287–307.

Harker, D. (1956). *The determination of the phases of the structure factors of non-centrosymmetric crystals by the method of double isomorphous replacement. Acta Cryst.* **9**, 1–9.

Hendrickson, W. A., Horton, J. R. & LeMaster, D. M. (1990). *Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. EMBO J.* **9**, 1665–1672.

Löwe, J., Stock, D., Jap, B., Zwickl, P., Baumeister, W. & Huber, R. (1995). *Crystal structure of the 20S proteosome from the archaeon T. acidophilum at 3.4 Å resolution. Science*, **268**, 533–539.

McRee, D. E. (1998). *Practical protein crystallography.* San Diego: Academic Press.