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

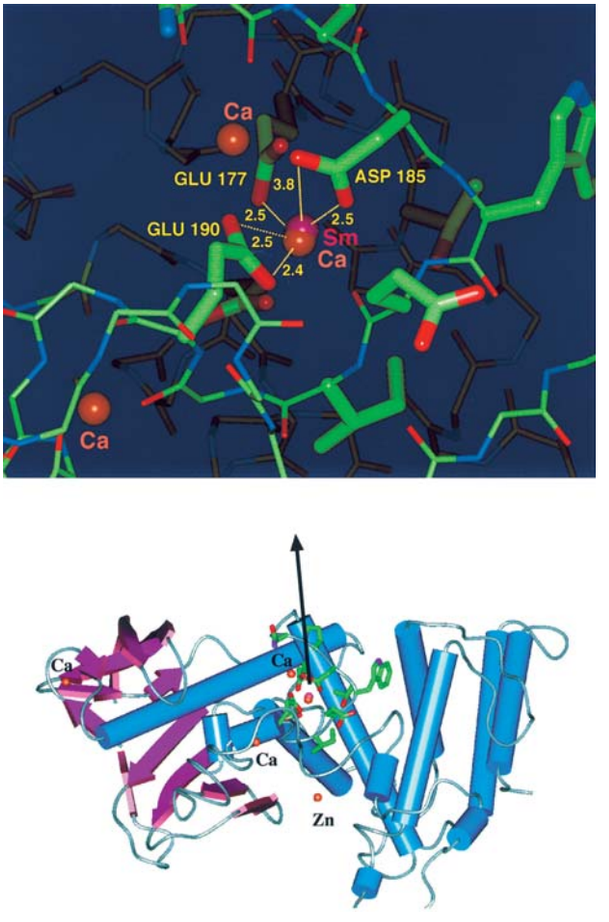

Fig. 12.1.6.1. The displacement of calcium by samarium in thermolysin. The samarium of the heavy-atom derivative is shown superposed on the parent crystal structure.

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.