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.

references