# [13] Preparation of Isomorphous Heavy-Atom Derivatives

*By* GREGORY A. PETSKO

This chapter will concentrate on the chemical logic underlying the preparation of heavy-atom derivatives of proteins. Most previous discussions[1] have tabulated those compounds and reaction conditions that have produced useful derivatives in various crystalline proteins. The present treatment follows the pioneering discussion of Blake,[2] and attempts to set down a systematic approach that can be used on future problems. First, the various amino acid side chains are classified according to their reactivity with different heavy metals. Then, different classes of heavy atoms are discussed, followed by methods for the introduction of reactive functionalities into proteins. The reader is also encouraged to consult the chapter on heavy-atom derivative preparation in Blundell and Johnson.[3]

## General Considerations

Whenever a new crystalline protein must be derivatized for phase determination by the method of isomorphous replacement, the crystallographer faces an initial choice: are the derivatives to be prepared by reacting the protein with heavy metal in solution and then crystallizing the complex, or by soaking the crystals in mother liquor containing the heavy-atom compound? Historically, both methods have been employed with success, but this author believes that diffusion into preexisting native crystals should be tried first. Intermolecular contacts in the crystal lattice will reduce the number of potentially reactive groups available for heavy-atom binding, making it more likely that the important first derivative will have a small number of sites. Also, use of the native crystal form eliminates the problem of finding crystallization conditions for a derivatized protein whose solubility may have changed, and greatly reduces the possibility that the derivative crystal will not be isomorphous with the parent form. If the soaking method does not work and the protein seems unreactive, exposure to heavy-atom compounds in solution followed by crystallization should then be tried, as the restrictive effects of lattice contacts will be avoided.

[1] D. Eisenberg, *in* "The Enzymes" (P. D. Boyer, ed.), 3rd ed., vol. 1, p. 1. Academic Press, New York, 1970.

[2] C. C. F. Blake, *Adv. Protein Chem.* **23**, 59 (1968).

[3] T. L. Blundell and L. N. Johnson, "Protein Crystallography." Academic Press, New York, 1976.

Sigler and Blow[4] noted the possible competition of $NH_3$ derived from $(NH_4)_2SO_4$ for heavy-metal binding to the protein, and it is clear that the preparation of derivatives will depend on the pH, composition of the mother liquor, and temperature. Specific effects will be noted below, but in general the following rules apply.

1. Ammonium sulfate, except at low pH (i.e., below 6), is a poor mother liquor for heavy-atom binding due to the production of the good nucleophile $NH_3$. When possible, ammonium sulfate should be replaced by sodium or lithium sulfate, or potassium phosphate. However, phosphate is a poor mother liquor for uranium and rare earth binding.
2. Intermediate pH (i.e., 6–8) is better than high pH (>9) because many heavy-atom compounds are alkalai labile or form insoluble hydroxides, but pH values below 6 may cause problems because most reactive groups will be protonated and blocked at low pH.
3. Most heavy-atom binding will be significantly slower at 4° than at room temperature.
4. Tris buffer is a potential competitor with the protein for heavy metals such as platinum, while phosphate buffer is a potential competitor for uranyl and rare earth compounds.

Soaking time and heavy-atom concentration are the other remaining considerations. Published soaking times are as short as several hours or as long as months. If no binding is found in a soak of several days, increasing the soaking time to 1 or more weeks may produce some binding, but cases of this are rare. Generally, a soak of 1–3 days will suffice to screen for binding. The concentration of the heavy-metal compound will depend to some extent on its solubility in the crystal mother liquor, but 1 m*M* seems a useful first value. Concentration is a more useful variable than soaking time if no binding is found, since mass action can be used to produce reasonable occupancy of even weak complexes. Increasing the heavy-atom concentration by an order of magnitude, or more, is worthwhile. If excessive binding is observed, by either complex Patterson maps, huge intensity changes, or crystal cracking, reduction of either soaking time or heavy-atom concentration, or both, is warranted. The author has found that, particularly in cases of covalent binding, both the time and the concentration can be very small. Soaking times as short as 1 hr combined with concentrations of 0.01 m*M* have produced full occupancy mercury derivatives of proteins with reactive sulfhydryl groups.[5]

[4] P. B. Sigler and D. M. Blow, *J. Mol. Biol.* **14**, 640 (1965).
[5] D. Ringe, G. A. Petsko, F. Yamakura, K. Suzuki, and D. Ohmori, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 3879 (1983).

Two final comments concern the use of cross-linking agents and the action of light. If the heavy-metal binding causes severe crystal damage, even at low concentrations with brief incubation, prior cross-linking of the crystal with glutaraldehyde may prevent cracking. Finally, many heavy-atom compounds, particularly platinum and iridium complexes, have a very vigorous photochemistry. Reproducibility of derivative occupancy—and, sometimes, of sites—can best be ensured by carrying out all experiments in the dark.

## Reactivities of the Amino Acids

Often the choice of the first heavy-atom compounds to try, and the appropriate conditions of pH and mother liquor composition, can be made from inspection of the amino acid composition of the protein. The following discussion of the reactivities of the 20 commonly occurring amino acids is meant to guide such choices.

Nearly all heavy-atoms react as cations or as complexes by nucleophilic ligand substitution reactions (the chief exception is iodine, which reacts with activated aromatic residues by electrophilic substitution). Therefore, the reactive amino acids can be discussed with reference to their nucleophilicity and their ability to substitute for common heavy-atom ligands.

*Unreactive Amino Acids.* The side chains of glycine, alanine, valine, leucine, isoleucine, proline, and phenylalanine are not reactive under the chemical conditions possible for crystal soaking experiments.

*Cysteine.* Cysteine is highly reactive as the thiolate anion, the percentage of which increases rapidly above pH 7. Sulfur in the ionic form is an excellent nucleophile and will react rapidly and essentially irreversibly with mercuric ion or organomercurials. In ligand substitution reactions, the thiolate is a fast-entering attacking group that forms thermodynamically stable complexes with class b metals (copper, silver, iridium, rhodium, platinum, paladium, gold). Reaction with rare earths and uranyl complexes is not expected. The substitution reactions are more sensitive to pH than the covalent attack. At pH values of 6 and below, where cysteine is nearly completely protonated, reaction with mercurials is still rapid but there is little reactivity with $PtCl_4^{2-}$.[6]

*Cystine.* Surprisingly, disulfide linkages are weakly reactive in ligand substitution reactions. $PtCl_4^{2-}$ has been found to bind to S–S bridges in some crystalline proteins by loss of chloride. Mercurials do not ordinarily

[6] G. A. Petsko, D. C. Phillips, R. J. P. Williams, and I. A. Wilson, *J. Mol. Biol.* **120,** 345 (1978).

insert spontaneously into disulfides but addition of mercury is possible if the cystine is first reduced with, say, dithiothreitol prior to mercuration.[7] Of course, steric restrictions may prevent formation of the S–Hg–S group.

*Histidine.* Histidine is the most common reactive group in proteins. The protonated imidazolium cation predominates below pH 6 and is not reactive as a nucleophile. The unprotonated imidazole is a good nucleophile and will displace chloride (or hydroxide) from platinum complexes, coordinate to gold complexes, react with copper and silver, and can be mercurated. The best heavy atoms for reaction with histidine are $K_2PtCl_4$, ethyl mercury phosphate, mersalyl, and $NaAuCl_4$. The imidazole ring is aromatic and electrophilic aromatic substitution reactions, such as iodination, are possible for histidine but the conditions are somewhat severe.

*Methionine.* The un-ionizable $S–CH_3$ group is unreactive toward mercurials but its lone pair of electrons makes it a good attacking group in nucleophilic ligand substitution reactions. Methionine will readily displace chloride from all platinum chloride complexes to form a stable bond. Contrary to the earlier hypothesis of Dickerson and associates,[8] this reaction does not involve oxidation of the platinum to the +4 state. It is a simple, $S_N2$-type ligand exchange reaction.[6] The reaction of platinum compounds with methionine is not pH sensitive in the normal range.

*Lysine.* The amino group of lysine has a $pK_a$ of about 9 in proteins. At neutral pH or lower, the protonated amine will not participate in substitution reactions but may undergo weak electrostatic complexation with negatively charged heavy-metal compounds. Near and above the $pK_a$, lysine can react with platinum and gold complexes to displace weaker ligands such as chloride (iodine, bromine, and nitro groups will not be displaced by lysine but may be by histidine and should be by the sulfur-containing amino acids). The particular utility of lysine residues, which are often highly accessible to the solvent, is their use as "handles" to attach other, more reactive, functionalities to proteins (see below).

*Arginine.* Arginine would seem to be a very attractive group for derivatization as it occurs relatively infrequently in most proteins and, therefore, would provide derivatives with a small number of sites per molecule. Unfortunately, the $pK_a$ of the guanidinium group is >12 in proteins and the cationic form is unreactive in simple substitution reactions. There may be an electrostatic interaction with anionic heavy-atom compounds.

[7] K. R. Ely, R. L. Girling, M. Schiffer, D. E. Cunningham, and A. B. Edmundson, *Biochemistry* **12**, 4233 (1973).

[8] R. E. Dickerson, D. Eisenberg, J. Varnum, and M. L. Kopka, *J. Mol. Biol.* **45**, 77 (1969).

Recently, advantage has been taken of the facile reaction of guanidinium groups with the glyoxal function (phenylglyoxal is a common arginine reagent in protein chemical modification) by using mercury phenylglyoxal as an arginine-specific heavy-atom compound.[9] This reagent is a useful addition to the crystallographic arsenal for protein crystals at pH values of neutral or greater.

*Tryptophan.* The indole ring is relatively inert to electrophilic aromatic attack by iodine but the ring nitrogen can be mercurated.[10] The pH is not critical for this reaction, but there must be no competing nucleophiles in the mother liquor. Optimal conditions are sodium sulfate or polyethylene glycol as the precipitant with cacodylate or borate buffer. The mercurial reagents that have proven successful are ethylmercury phosphate and mercuric acetate. Tryptophan does not react readily in ligand substitution processes.

*Tyrosine.* The phenolate oxygen anion is a very good nucleophile and would be expected to bind a variety of heavy atoms if not for its high $pK_a$ (10.5). Only at pH values >10 would there be sufficient concentration of the phenolate form. Therefore, the principal reaction of tyrosine is electrophilic aromatic substitution by iodine. Sigler[11] has given a detailed description of the preparation of the iodination solution and the conditions for its use with crystalline proteins. It is worth remembering that ammonium sulfate mother liquor at high pH should be avoided if iodination is attempted: one may accidently synthesize nitrogen triiodide, which is explosive!

*Aspartic and Glutamic Acids.* Both of these residues have side chain $pK_a$ values on the order of 3–4. At low pH they will be protonated and unreactive. Above pH 5 they will be anionic and potentially good ligands for heavy metals. Although class b metals prefer sulfur or nitrogen ligands, uranium and various rare earths behave like calcium and display a strong preference for oxygen ligation. Colman *et al.*[12] have shown that lanthanides can replace calcium in thermolysin; the rare earths of higher atomic number, having ionic radii less than or equal to that of calcium, were found to cause less perturbation of the protein structure. In all cases, the primary ligands are carboxylate side chains from aspartic and glutamic acids. Uranyl complexes such as $K_3UO_2F_5$ and $UO_2(NO_3)_2$ are also carboxylate seeking, but these compounds are not very soluble above pH 7 due to the formation of hydroxides. Uranium and rare earth derivati-

[9] I. A. Wilson, J. J. Skeral, and D. C. Wiley, *Nature (London)* **289,** 366 (1981).
[10] D. Tsernoglou and G. A. Petsko, *FEBS Lett.* **68,** 1 (1976).
[11] P. B. Sigler, *Biochemistry* **9,** 3609 (1970).
[12] P. M. Colman, L. H. Weaver, and B. W. Matthews, *Biochem. Biophys. Res. Commun.* **46,** 1999 (1972).

zation must not be attempted in phosphate buffer, because the phosphate oxyens will compete with protein carboxylates for the heavy atom, and the phosphate complexes are often insoluble.

*Serine and Threonine.* The hydroxyl groups of these side chains are fully protonated at all normal pH values and are not reactive nucleophiles. Occasionally, an abnormally reactive serine has been found as in the serine proteases and $\beta$-lactamases, and these residues have been derivatized, but such behavior is uncommon.

*Asparagine and Glutamine.* In model studies with simple metal complexes the free amino acids Asn and Gln will weakly coordinate through their amide nitrogens, but such complexes have not been found frequently in crystalline proteins. At present there is no information that permits a rational approach to the labeling of these amino acids with heavy metals. A systematic study would be worthwhile.

## Classes of Heavy Atoms

*Platinum Compounds.* The most widely successful heavy-atom reagent is the $PtCl_4^{2-}$ ion, which binds to methionine, histidine, and cysteine residues. In the absence of excess chloride ion this compound is rapidly converted to aquo and hydroxide species of neutral or +1 overall charge. Petsko *et al.*[6] gave a detailed account of the chemistry of chloroplatinate in a variety of crystal mother liquors. Briefly, they concluded that ammonium sulfate solutions at high pH will compete with the protein for platinum binding, that pH values greater than 5–6 are needed to unblock histidine so that it will react with chloroplatinate, and that cysteine is a ligand only at pH values near neutrality or above. They concluded further that most other platinum compounds, such as $Pt(NO_2)_4^{2-}$, $Pt(NH_3)_2Cl_2$, or Pt(ethylenediamine)$Cl_2$ will react in the same way. The exceptions are compounds with ligands of high thermodynamic stability, such as $Pt(CN)_4^{2-}$. The cyanide will not be displaced by any protein side chain, so platinum tetracyanide is inert to nucleophilic ligand substitution. This complex remains dianionic under all conditions, and will bind, if at all, to positively charged protein residues (lysine, arginine, and, at low pH, histidine) by electrostatic attraction.

*Mercurials.* A wide range of mercury compounds have been observed to bind to crystalline proteins. With only a few exceptions, the protein ligands are either cysteine sulfur or histidine nitrogen. Binding is promoted at pH values of neutral or greater, where these residues are deprotonated. The most commonly used mercurials are *p*-chloromercuribenzoic acid, ethylmercury thiosalicylate, $HgCl_2$, ethylmercury chloride, mercuric acetate, mersalyl, and ethylmercury phosphate. The latter com-

pound is the author's personal favorite. It is commercially available (Chem Service), water soluble (except in the presence of phosphate ions; phosphate buffer should never be used with this reagent), and disproportionates to give the highly reactive, partially hydrophobic $CH_3CH_2Hg^+$ ion.[13] This species will penetrate into proteins to react with buried sulfhydryl or imidazole groups and, being small, usually does not cause lack of isomorphism. If too many sites are found by this reagent, greater steric selectivity may be achieved with a larger, aromatic mercurial. Once even a single site has been substituted with mercury, a second isomorphous derivative often may be prepared by repeating the experiment with a dimercurial. Dimercury acetate is a popular choice; the two mercury atoms are 3.8 Å apart and the compound is not very large. Sprang and Fletterick[14] gave a simple synthesis for this dimercurial. Although competitive binding of $NH_3$ from the mother liquor is less of a problem for mercurials than it is for platinum compounds, the author's recommendation is still to replace ammonium sulfate with lithium or sodium sulfate if no binding is found.

*Rare Earths and Uranyl Salts.* As stated above, phosphate buffer will tend to compete with protein carboxylate groups for binding to lanthanides or uranium compounds, so this buffer should be replaced by Tris, cacodylate, or borate (depending on the pH range). Samarium is the preferred lanthanide owing to its large anomalous scattering signal. Of the commonly available uranyl salts, the nitrate is the most reactive but the acetate is slightly more soluble in common mother liquors.

*Complex Ions.* One species stands out in this category: $K_2HgI_4$, which disproportionates in aqueous solution to give a mixture of substances: $HgI_4^{2-}$, $HgI_3^-$, and $HgI_2$ as well as $I^-$. The most important product is the planar, trigonal $HgI_3^-$ group, which will bind electrostatically to cations or penetrate into proteins by virtue of its large flat structure. The dissociation equilibrium for this ion can be driven in the direction of $HgI_3^-$ formation by the addition of excess KI to the $K_2HgI_4$ soaking solution.

## Special Reactions

One potentially reactive nucleophile is present in all proteins, the terminal amino group (the C-terminal carboxylate does not appear to bind heavy metals). This group has been exploited by Drenth and associates,[15] who have reacted the amino terminus of papain with an iodinated Ed-

[13] D. A. Vidusek, M. F. Roberts, and G. Bodenhausen, *J. Am. Chem. Soc.* **104,** 5452 (1982).
[14] S. Sprang and R. J. Fletterick, *J. Mol. Biol.* **131,** 523 (1979).
[15] J. Drenth, J. N. Jansonius, R. Koekoeck, H. M. Swen, and B. G. Wolthers, *Nature (London)* **218,** 929 (1968).

man's reagent. This reaction has been somewhat neglected, in the author's opinion unjustly. Even if an iodine is not heavy enough to serve as a major derivative for a protein, location of the N-terminus would be a valuable aid in chain tracing. There are two complications in the use of Edman-like compounds: protonation of the amino terminus and competition from lysine reaction. As usual, it is the unprotonated amine that is the reactive nucleophile so exposure to the Edman compound should be done above neutral pH. Unfortunately, the unprotonated $\epsilon$-amino group of lysine will also react with Edman reagents. Control of the reaction may be achieved sometimes by taking advantage of the lower $pK_a$ of the protein N-terminus. Carrying out the reaction between pH 7 and 8 will sometimes give only substitution at the amino terminus. Of course, $NH_3$ derived from the ammonium ion will rapidly inactivate Edman compounds, so ammonium sulfate mother liquor cannot be used.

Another method of specific chemical labeling of crystalline enzymes is by heavy-atom-substituted substrate analogs or inhibitors. Unfortunately, each protein presents a different case, and no general strategy can be given. However, it is relatively easy to mercurate or iodinate aromatic compounds, so synthesis of suitable analogs is often possible.

Some proteins in their native state bind metal atoms that can be replaced by heavier atoms, thus providing an isomorphous derivative. The metal most often replaced in such studies has been zinc. Removal of zinc is achieved by dialysis against EDTA, 1,10-phenanthroline, histidine, or some combination of these chelators. Then cadmium or mercury or lead (usually as the acetate salt) is added by dialysis or direct soaking. Lanthanides will replace bound calcium atoms in many proteins on simple soaking of the crystals in solutions containing the rare earth.

### Insoluble Reagents

Many heavy-atom-containing compounds, such as derivatives of Edman's reagent, are at best only sparingly soluble in water. Addition of these reagents to protein crystals can be achieved by use of a carrier solvent that is less polar. Acetonitrile is the most popular choice. The heavy-atom compound should be dissolved in neat, vacuum-distilled acetonitrile and the organic solvent should then be suspended in the crystal mother liquor at 3–5% by volume.

### Introduction of Reactive Groups

When all of the standard, trial-and-error methods for preparation of a heavy-atom derivative have failed, the last resort is to introduce one or

more highly reactive functionalities into the protein, and then to react the modified macromolecule with heavy metal. Two strategies are possible: the protein can be modified in solution and then crystallized, or the chemical reactions can be carried out in the crystalline state. The author prefers to start with the latter approach, as the crystal contacts should limit the number of modification sites and reduce the chances for nonisomorphism.

Because of their high degree of reactivity to mercurials, sulfhydryl groups are the logical choice for introduction into proteins. The usual target group for the modification reaction is the amino group of lysine (or the protein N-terminus). Shall and Barnard reacted ribonuclease with *N*-acetylhomocysteine thiolactone and produced two separable products, each containing a new SH group in a different place.[16] This reaction was run in solution, but it should be possible to carry it out in the crystalline state at neutral pH. Gallwitz *et al.*[17] reacted tobacco mosaic virus with 4-sulfophenylisothiocyanate, a modified Edman-like reagent, which derivatized both the N-terminus and lysine 53 of the coat protein. The reaction yielded phenylthiocarbamoyl derivatives that tautomerized to give reactive SH groups. The sulfhydryls reacted with methylmercury nitrate to give a stable thiomercurial that did not suffer replacement of the sulfur by oxygen in air. Gallwitz *et al.*[17] gave instructions for the synthesis of the modification reagent. Reaction of this compound with proteins must be carried out in the dark anaerobically at pH 8.5.

Mowbray and Petsko[18] have described the reaction of amino groups in the galactose chemotaxis receptor protein with carbon disulfide under anaerobic conditions in the crystalline state. They described a simple apparatus for carrying out the modification, in phosphate buffer at pH 8. Subsequent reaction with mercurials and dimercurials produced several different isomorphous derivatives for a crystalline protein that had not been successfully derivatized by conventional methods. The carbon disulfide reaction appears to form a dithiocarbamate with a $pK_a$ of 3.5. The modified amino group is only stable above pH 6, so the mercuration must be carried out at neutral pH or higher.

Finally, metal-chelating groups can be attached to protein amino groups by similar chemistry. Benisek and Richards[19] reacted lysozyme with picolinimidate compounds, with up to seven such groups being covalently attached to the protein, depending on conditions. These groups chelated platinum and gold compounds very strongly.

[16] S. Shall and E. A. Barnard, *J. Mol. Biol.* **41**, 237 (1969).
[17] U. Gallwitz, L. King, and R. N. Perham, *J. Mol. Biol.* **87**, 257 (1974).
[18] S. L. Mowbray and G. A. Petsko, *J. Biol. Chem.* **258**, 5634 (1983).
[19] W. Benisek and F. M. Richards, *J. Biol. Chem.* **243**, 4267 (1968).

Recommendations

Consideration of the specific amino acid composition, presence of endogenous metals, cysteine content, etc. is always desirable; nevertheless, some general rules can be formulated.

1. Whenever possible, avoid using ammonium sulfate unless the pH is below 6.
2. Heavy-metal binding is more likely at pH 7 and above than at pH values below 6.
3. All soaking should be done in the dark.
4. For first screening of possible derivatives, the author recommends (a) $K_2PtCl_4$ at pH 6 or higher in ammonia-free media; (b) ethylmercury phosphate in phosphate-free media, pH 6 or higher; (c) samarium acetate or uranyl acetate in phosphate-free media; (d) $K_2HgI_4$ with excess KI; or (e) $K_2Pt(CN)_4$.

It would be surprising (but, alas, not unprecedented) if none of these five trials produced significant intensity changes. In the author's opinion, to produce *some* binding to any given protein, the reagent of choice would be $K_2PtCl_4$ or the corresponding nitrite. To produce a *suitable* isomorphous derivative of high occupancy with a reasonable number of sites, the author would suggest ethyl mercury phosphate.

*Cautionary Postscript*

*All* heavy-atom compounds should be regarded as highly toxic. Mouth pipetting of heavy-atom-containing solutions should never be attempted. Preparation of such solutions must be done in a fume hood. Gloves and protective clothing and eyeguards should be worn at all times.