Trimethyllead Acetate: A First-Choice Heavy Atom Derivative for Protein Crystallography

Hazel M. Holden* and Ivan Rayment†

Institute for Enzyme Research, Graduate School, and *Department of Chemistry, College of Letters and Science and †Department of Biochemistry, College of Agricultural and Life Sciences, University of Wisconsin, Madison, Wisconsin 53705

Received June 17, 1991, and in revised form August 13, 1991

The three-dimensional conformation of a protein provides a wealth of biochemical information and with the advent of cloning techniques that allow the preparation of proteins almost at will, a renewed interest has arisen in the crystallographic determination of protein structures. As in any research technique, however, there are often many difficulties encountered in an X-ray crystallographic investigation. One of these is the "phase problem." Although in recent years there has been considerable progress in the development of techniques for phase determination, including the use of molecular replacement and multiple wavelength measurements, the multiple isomorphous replacement method is still the most successful method for obtaining a three-dimensional structure. Here we report the use of trimethyllead acetate as a heavy atom compound of first choice in the preparation of an isomorphous heavy atom derivative.

Multiple isomorphous replacement, which involves the preparation of heavy atom derivatives, was the first technique successfully employed to overcome the "phase problem" in protein crystallography during the structural analysis of myoglobin and has since enjoyed widespread application (1, 2). The preparation of suitable isomorphous heavy atom derivatives requires that heavy metals bind to the protein in such a way as to not disrupt the crystalline lattice or to effect major protein conformational changes. Over the years a wide range of compounds containing mercury, platinum, gold, osmium, and uranium, to name but a few, has been employed in various X-ray analyses of proteins. These reagents exhibit various degrees of specificity toward a limited number of amino acid side chains. The reactivities of these compounds are often controlled by the composition of the crystallization medium and mediated by the immediate surrounding environment of the side chain to which the heavy atom is binding. Preparation of an isomorphous heavy atom derivative usually represents a compromise of attaining a satisfactory level of substitution at a hopefully restricted number of binding sites to the protein without disrupting the crystalline lattice. This compromise is generally achieved by variation of the reagent concentration, the composition of the crystallization medium, and the soaking time.

Several problems frequently encountered in the preparation of heavy atom derivatives include insolubility of the reagent in the crystallization medium, reactivity of the heavy metal with constituents of the buffer and precipitant, and too aggressive reaction of the heavy atom compound with the protein, leading to multiple, low-occupancy heavy atom binding sites. Consequently, an ideal derivative would be one that is soluble in all commonly employed precipitants and demonstrates a high specificity for a limited number of binding sites on the protein. As shown by the eight structural analyses described here, insecticyanin from Manduca sexta L., cytochrome c2 from Rhodobacter capsulatus, cytochrome c2 from Paracoccus denitrificans, apolipoprotein-III from Locusta migratoria, the [2Fe–2S] ferredoxin from Anabaena, the high-potential iron–sulfur protein from Ectothiorhodospira halophila, UDP-Gal 4-epimerase from Escherichia coli, and myosin subfragment-1, trimethyllead acetate fulfills these requirements. It is especially noteworthy that the abovementioned proteins represent a wide range of crystallization conditions and molecular weights ranging from the smallest of 8000 (E. halophila high-potential iron–sulfur protein) to the largest of 130,000 (myosin subfragment-1).
1. Relevant crystal statistics for each of these proteins may be found in Table I.

**MATERIALS AND METHODS**

Trimethyllead acetate was purchased from Aldrich Chemical Co., Milwaukee, Wisconsin. Ammonium sulfate was enzyme grade and obtained from Schwarz/Mann. Polyethylene glycol was purchased from either Sigma Chemical Co., or Aldrich. All buffers were obtained from Research Organics, Cleveland, Ohio, and all other chemicals were reagent grade and purchased through VWR or Sigma.

The lead derivative X-ray data sets were scaled to the respective native X-ray data in shells of equal volume in reciprocal space based on resolution. Scaling statistics may be found in Table II. The positions, occupancies, and thermal parameters of the lead binding sites in the crystalline proteins were refined with the origin-removed Patterson-function correlation method (3) using the program HEAVY (4). Phasing statistics, as determined from HEAVY, are presented in Table III.

**RESULTS AND DISCUSSION**

Our first experience with trimethyllead acetate as a heavy atom derivative was in the structure determination of insecticyanin, a blue bilirubin isolated from the hemolymph of the tobacco hornworm *M. sexta* L. and believed to play a key role in the cryptic coloration of the insect (5). Insecticyanin is a tetramer comprising identical subunits, each containing 189 amino acid residues (6). Initially, heavy atom derivative searches were conducted at pH 5.5 but all attempts to prepare such derivatives were unsuccessful. It was found, however, that the native crystals could be transferred to a synthetic mother liquid buffered at pH 7.5 with 50 mM Na⁺/K⁺ phosphate with no apparent deterioration of their diffraction properties or changes in their physical morphology. At this pH, three heavy atom derivatives were readily prepared using 10 mM trimethyllead acetate, 5 mM *p*-hydroxymercuriphenyl sulfonic acid, and 5 mM K₂PtCl₆ (8). The lead derivative proved to be highly isomorphous, producing uniform differences between the native and derivative structure factors across the entire resolution range with only very small changes in the unit cell dimensions (Table II).

The locations of the heavy atom binding sites for the lead derivative were determined by visual inspection of a difference Patterson map calculated with X-ray data from 30 to 5 Å. All of the Harker peaks and cross vectors for the two heavy atom sites per asymmetric unit could be

### Table I

<table>
<thead>
<tr>
<th>Protein</th>
<th>Space group</th>
<th>Unit cell dimensions (Å)</th>
<th>Crystallization medium</th>
<th>No. of molecules per asymmetric unit</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Insecticyanin</td>
<td><em>P</em>₄₂₁₂</td>
<td><em>a</em> = 115.4, <em>c</em> = 71.5</td>
<td>20% polyethylene glycol 8000, 100 mM succinate, 100 mM NaCl, 5 mM Na₂SO₄, pH 5.5, room temperature</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>Apolipophorin-III</td>
<td><em>P</em>₆₂₂₂</td>
<td><em>a</em> = 67.5, <em>c</em> = 155.6</td>
<td>2.8 M ammonium sulfate, 50 mM Na⁺/K⁺ phosphate, 5 mM NaNO₃, pH 7.4, room temperature</td>
<td>1</td>
<td>12</td>
</tr>
<tr>
<td>Cytochrome <em>c₁</em> from <em>Rb. capsulatus</em></td>
<td><em>R</em>32</td>
<td><em>a</em> = 100.0, <em>c</em> = 162.2</td>
<td>3.2 to 3.5 M ammonium sulfate, 250 mM NaCl, 50 mM NaNO₃, pH 7.5, room temperature</td>
<td>2</td>
<td>14</td>
</tr>
<tr>
<td>Cytochrome <em>c₂</em> from <em>P. denitrificans</em></td>
<td><em>P</em>₂₁₂₂₁</td>
<td><em>a</em> = 31.6, <em>b</em> = 42.1, <em>c</em> = 81.6</td>
<td>3.4 M ammonium sulfate, 50 mM HEPES, 5 mM Na₂SO₄, pH 7.5, room temperature</td>
<td>1</td>
<td>16, 17</td>
</tr>
<tr>
<td>High-potential iron-sulfur protein from <em>E. halophila</em></td>
<td><em>P</em>₂₁</td>
<td><em>a</em> = 60.0, <em>b</em> = 31.9, <em>c</em> = 40.3, <em>β</em> = 100.5</td>
<td>3.4 to 3.5 M ammonium sulfate, 50 mM Na⁺/K⁺ phosphate, 5 mM Na₂SO₄, pH 7.0, room temperature</td>
<td>2</td>
<td>18</td>
</tr>
<tr>
<td>[2Fe-2S] Ferredoxin from <em>Anabaena</em> 7120</td>
<td><em>P</em>₂₁₂₂₁</td>
<td><em>a</em> = 37.6, <em>b</em> = 38.1, <em>c</em> = 147.3</td>
<td>2.6 M ammonium sulfate, 50 mM succinate, 1% 2-methyl-2,4-pentanediol, 5 mM Na₂SO₄, pH 5.5, 4°C</td>
<td>2</td>
<td>20</td>
</tr>
<tr>
<td>Myosin subfragment-1</td>
<td><em>C</em>₂₂₂₁</td>
<td><em>a</em> = 98.4, <em>b</em> = 124.9, <em>c</em> = 275</td>
<td>1.5 M ammonium sulfate, 500 mM KCl, pH 6.5, 4°C</td>
<td>1</td>
<td>21</td>
</tr>
<tr>
<td>UDP-Gal 4-epimerase from <em>E. coli</em></td>
<td><em>P</em>₂₁₂₂₁</td>
<td><em>a</em> = 76.3, <em>b</em> = 83.1, <em>c</em> = 132.1</td>
<td>6.5% polyethylene glycol 8000, 50 mM succinate, 200 mM NaCl, 5 mM Na₂SO₄, pH 6.0, 4°C</td>
<td>1</td>
<td>23</td>
</tr>
</tbody>
</table>
TABLE II
Scaling Statistics between Native and Derivative Crystals

<table>
<thead>
<tr>
<th>Protein</th>
<th>Resolution range (Å)</th>
<th>Average isomorphous differences (%) a</th>
</tr>
</thead>
<tbody>
<tr>
<td>Insecticyanin</td>
<td>5 to 7.1</td>
<td>15.9</td>
</tr>
<tr>
<td>Apolipoporin-III</td>
<td>4.6 to 5.60</td>
<td>24.6</td>
</tr>
<tr>
<td><em>Rh. capsulatus</em> cytochrome c₂</td>
<td>2.3 to 2.58</td>
<td>24.9</td>
</tr>
<tr>
<td><em>P. denitrificans</em> cytochrome c₂</td>
<td>2.5 to 3.17</td>
<td>23.5</td>
</tr>
<tr>
<td><em>E. halophilia</em> high-potential iron-sulfur protein</td>
<td>2.0 to 2.54</td>
<td>19.6</td>
</tr>
<tr>
<td><em>Anabaena</em> [2Fe-2S] ferredoxin</td>
<td>1.9 to 2.54</td>
<td>17.4</td>
</tr>
<tr>
<td>Myosin subfragment-1</td>
<td>1.8 to 2.54</td>
<td>17.2</td>
</tr>
<tr>
<td>UDP-Gal 4-epimerase</td>
<td>1.6 to 2.54</td>
<td>15.8</td>
</tr>
</tbody>
</table>

a $R = \Sigma |F_N| - |F_H|/\Sigma |F_N|$, where $|F_N|$ is the native structure factor amplitude and $|F_H|$ is the lead derivative structure factor amplitude.

identified in the Patterson function. Most importantly, this derivative was used to locate the positions of the heavy atom binding sites for the other two derivatives by difference Fourier maps. It proved to be the best derivative in the structure determination as seen by the phasing statistics given in Table III. Once the molecular structure of insecticyanin had been determined and refined to 2.5 Å resolution, it became obvious, as shown by the electron density in Fig. 1, that the trimethyllead moiety had bound to the surface amino acid residue, Glu 68, in both subunits. Those amino acid residues within 5.5 Å of the lead binding site are shown in stereo in Fig. 2.

One of the major advantages of trimethyllead acetate as a heavy atom derivative is its tendency to bind at only a few locations as demonstrated with the insecticyanin structure. While there are 10 glutamate residues per subunit of insecticyanin, the lead compound substituted at only 1. This lack of multiple binding sites is especially valuable when attempting to deconvolute Patterson functions for high-symmetry space groups. Accordingly, trimethyllead acetate proved to be the critical derivative in the structure determination of an apolipoprotein, apolipoporin-III, isolated from the African migratory locust *L. migratoria* and the cytochrome c₂ from *Rh. capsulatus*,

TABLE III
Phase Calculation Statistics

<table>
<thead>
<tr>
<th>Protein</th>
<th>Resolution range (Å)</th>
<th>Phasing power for acentric reflections*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Insecticyanin</td>
<td>5.94</td>
<td>1.62</td>
</tr>
<tr>
<td>Apolipoporin-III</td>
<td>4.63</td>
<td>1.21</td>
</tr>
<tr>
<td><em>Rh. capsulatus</em> cytochrome c₂</td>
<td>3.92</td>
<td>2.18</td>
</tr>
<tr>
<td><em>P. denitrificans</em> cytochrome c₂</td>
<td>3.46</td>
<td>2.03</td>
</tr>
<tr>
<td><em>E. halophilia</em> high-potential iron-sulfur protein</td>
<td>3.13</td>
<td>1.31</td>
</tr>
<tr>
<td><em>Anabaena</em> [2Fe-2S] ferredoxin</td>
<td>2.88</td>
<td>1.53</td>
</tr>
<tr>
<td>Myosin subfragment-1</td>
<td>2.68</td>
<td>1.42</td>
</tr>
<tr>
<td>UDP-Gal 4-epimerase</td>
<td></td>
<td>1.46</td>
</tr>
</tbody>
</table>

*Phasing power is the ratio of the root-mean-square heavy atom scattering factor amplitude to the root-mean-square lack of closure error.
both of which crystallize in high-symmetry space groups (Table I).

Apolipoprophorin-III from L. migratoria is a monomeric apolipoprotein containing 161 amino acid residues and a total sugar content of approximately 12% (9–11). The three-dimensional architecture of apolipoprophorin-III was determined to 3.0 Å resolution by conventional multiple isomorphous replacement with five heavy atom derivatives (31). The lead bound at only one site in the asymmetric unit and gave a very clean Patterson function in which all the Harker peaks were readily located by inspection as shown in Fig. 3. Subsequently, this derivative was used to locate the heavy atom binding sites for four other derivatives by difference Fourier maps (13). As observed with insecticyanin, while there were 13 glutamate residues in apolipoporphorin-III, trimethyllead acetate bound only to Glu 10 at approximately 2.2 Å from the side chain carboxylate group.

The cytochrome c$_2$ from Rb. capsulatus contains 116 amino acid residues and one heme moiety. The three-dimensional structure of this c-type cytochrome was determined by multiple isomorphous replacement with a total of four heavy atom derivatives. Only the Patterson function for the lead derivative could be readily solved by inspection. The resulting protein phases based on this lead derivative were then used to solve the three other derivatives and hence the molecular structure (15). In the case of this protein, there was only one lead binding site and this site was not located near glutamate residues. Instead, the binding pocket for the heavy metal was located at the interface between the two unique cytochrome molecules in the asymmetric unit.

Since the cytochrome c$_2$ isolated from P. denitrificans shows a 50% primary sequence homology to the Rb. capsulatus cytochrome c$_2$, we were interested in comparing these proteins in detail. The original structure determination of the P. denitrificans molecule was reported by Timkovich and Dickerson (16). We have since refined the structure of this cytochrome to a nominal resolution of 1.7 Å (17). This crystalline protein also binds trimethyllead acetate. There were four major lead binding sites: one site wedged between the side chain carboxylates of Asp 93 and Glu 61, a second site located between the side chain carboxylates of Asp 92 and Glu 11, a third site situated between two glutamate residues, Glu 51 and Glu 73, and a fourth binding site lying next to the carboxylate side chain of Glu 9.

It is often difficult to prepare good heavy atom derivatives for very small proteins as was the case for the high-potential iron–sulfur protein isolated from E. halophila. This [4Fe–4S] ferredoxin contains only 71 amino acid residues, 8 of which are glutamates. All experiments designed to prepare isomorphous heavy atom derivatives with various platinum or mercury compounds failed. Fortunately, the trimethyllead acetate bound isomorphously and proved to be a superb derivative which was sufficient to determine the protein’s molecular structure (19). Precession photographs for the native and trimethyllead acetate derivative are shown in Fig. 4.

Iron–sulfur proteins often present a challenge in preparing isomorphous heavy atom derivatives due to the
reactivity of their metal cluster prosthetic groups toward many of the commonly employed reagents. This was especially true for the 98-amino-acid-residue [2Fe–2S] ferredoxin isolated from Anabaena 7120 which was recently solved in this laboratory (20). Attempts to prepare isomorphous heavy atom derivatives at pH 5.5 failed. However, by transferring the crystals to solutions containing 3.8 M ammonium sulfate, 1% 2-methyl-2,4-pentanediol buffered at pH 7.0 with NaCl/KCl phosphate, and 20 mM trimethyllead acetate, it was possible to prepare a suitable derivative. The crystals were soaked in the lead-containing solution for 14 days at 4°C. Inspection of the Patterson function revealed three lead binding sites with no evidence of any minor sites. This single derivative was of sufficient quality to solve the three-dimensional structure of the ferredoxin since it revealed the location of the iron–sulfur clusters. Once the positions of the two independent iron–sulfur clusters in the unit cell were known, it was possible to improve the quality of the electron density map by noncrystallographic symmetry averaging and solvent flattening. Two of the lead binding sites were located between the side chains of Glu 24 and Glu 32 in each ferredoxin molecule in the asymmetric unit. One of these lead binding sites is shown in Fig. 5. The third lead binding site was located between the interface of the two ferredoxins in the crystalline lattice.

The preparation of isomorphous heavy atom derivatives for large proteins presents a different type of problem in that there is a much greater probability of heavy atoms binding to many minor sites. In principle this may not be a drawback once the first derivative has been solved; the major problem, however, is finding the first solvable derivative. Trimethyllead acetate proved to be the critical derivative in the structural studies of myosin subfragment-1 which is the portion of the myosin molecule that is responsible for the generation of force through its interaction with actin and the concomitant hydrolysis of ATP. It is a large proteolytic fragment of molecular weight
130,000. Myosin subfragment-1 crystals were soaked in 20 mM trimethyllead acetate for 3 weeks. The difference Patterson function was readily interpretable at 4.5 Å resolution and revealed four heavy atom binding sites. There were no significant changes in the unit cell dimensions and damage to the diffraction properties of the crystals was minimal. It is important to note that the myosin subfragment-1 crystals were originally grown from ammonium phosphate which is a particularly undesirable precipitant for preparing isomorphous derivatives. Nonetheless, trimethyllead acetate still formed a derivative in the presence of ammonium phosphate and the heavy atom binding sites were nearly identical to those observed in ammonium sulfate. Unfortunately, it proved difficult to

FIG. 4. 13° precession photographs of the $h00$ zones for the native and lead containing *E. halophila* high-potential iron–sulfur protein crystals. The X-ray photographs were recorded with nickel-filtered copper $K\alpha$ radiation from a Rigaku RU 200 rotating anode X-ray generator operated at 40 kV and 50 mA with a 200-μm focal cup. The exposure time was typically 20 h at a crystal-to-film distance of 100 mm. (a) The $h00$ zone from a native crystal. (b) The $h00$ zone from a crystal soaked in 20 mM trimethyllead acetate for 6 days. As can be seen, there are many obvious differences between the native and the lead derivative diffraction photographs without large changes in the unit cell dimensions or deterioration of the diffraction properties.

FIG. 5. Close-up view of the lead binding site in the [2Fe-2S] ferredoxin. In the [2Fe-2S] ferredoxin from *Anabaena*, the lead binding site is between Glu 32 and Glu 24 and is within approximately 3.0 Å of each carboxylate side chain.
find other good heavy atom derivatives in the ammonium phosphate crystallization medium.

Trimethyllead acetate has also been used in the recent structure determination of UDP-Gal 4-epimerase from E. coli. This enzyme is involved in the conversion of galactose to glucose and, as isolated from E. coli, is a dimer of identical subunits with a total molecular weight of 79,000 (22). The difference Patterson map for the lead derivative was readily interpreted by inspection and showed there to be two heavy atom binding sites. The three-dimensional structure of epimerase was subsequently determined to 2.5 Å resolution with protein phases based on six heavy atom derivatives, including the lead (24), and demonstrated that the two lead binding sites bound to the same monomer in the asymmetric unit and were located near Glu 320 and Asp 271, respectively.

In addition to the crystals studied in our laboratory, trimethyllead acetate has proven to be a successful derivative for the structural investigations of rat intestinal fatty acid binding protein (25), P2 myelin protein (26), methionine aminopeptidase from E. coli (27), and the cytochrome c2 from Rb. sphaeroides (28). Concentrations of the lead derivative used for these structural analyses ranged anywhere from 5 to 40 mM in the case of the methionine aminopeptidase and the Rb. sphaeroides cytochrome c2, respectively.

Lead compounds have not received as much attention as potential heavy atom derivatives because of their limited solubility in many crystallization solutions. In particular, most lead salts of divalent and trivalent anions are insoluble which prevents their use with sulfate and phosphate solvents. Trimethyllead acetate functions as a large soft cation and as a consequence avoids the problems associated with most simple lead compounds. Triethyllead acetate has been used more often but unfortunately this compound is less specific than the trimethyl reagent perhaps due to its more hydrophilic nature. In addition, triethyllead acetate is less soluble and in our experience generally causes more damage to the protein crystals.

Examination of the binding sites for trimethyllead acetate in the proteins described above reveals that there are two modes of binding. In one of these, the ion sits in a small hydrophobic pocket either within the protein or formed by protein-protein contacts within the crystal. In the other, the ion is coordinated by two carboxylate groups that lie in close proximity. In general the occupancy of the latter is higher than the former and occurs more frequently in the structures studied thus far.

The small molecule crystal structure of trimethyllead acetate reveals that the stereochemistry of the trimethyllead cation is not tetrahedral as might be expected, but rather is a distorted trigonal bipyramidal arrangement of ligands in which there are two axial oxygen ligands from different carboxyl groups (29). These groups form poly-

meric chains within the crystalline lattice. The distribution of ligands observed in the protein complexes reflects the arrangement seen in the small molecule structure. Even though at the resolution of the protein electron density maps it is not possible to unequivocally define the ligands involved in coordinating the trimethyllead cation or to define the positions of the methyl groups attached to the lead atom, it is clear that in the preferred binding mode the ion is coordinated by two charged carboxyl groups. At the resolution of the protein X-ray data it is difficult, however, to decide whether one or both oxygen atoms from the carboxylate groups are coordinated to the lead atom.

It appears that trimethyllead acetate functions best when the pH of the crystals is greater than 6.5 which reflects its preference for binding between two carboxylate groups that are in close proximity. This type of coordination would require that both groups be in the ionized state. The pKₐ for an isolated glutamate or aspartate amino acid falls around 4.5. However, it can be expected that the dissociation constant of two carboxyl groups in close proximity to one another will be perturbed from this value such that a higher pH will be required to ensure that both groups are mostly ionized.

In conclusion, trimethyllead acetate has proven to be a superb heavy atom derivative for at least 12 protein structural investigations within the past 5 years. It seems to work best at high concentrations, anywhere from 20 to 40 mM, and when the crystallization medium is above pH 6.5. Also, it is important to be patient. Many of the crystals were soaked in the lead derivative for well over a week and, in the case of insecticin, it required a 3-week soak. Clearly, this compound should be seriously considered when a search for suitable isomorphous heavy atom derivatives is initiated.

REFERENCES
27. Personal communication, Dr. Steven L. Roderick, University of Oregon, Eugene.
28. Personal communication, Dr. Herb Axelrod, University of California, San Diego.