Published in final edited form as:

FEBS J. 2014 September ; 281(18): 4021–4028. doi:10.1111/febs.12910.

A rapid and rational approach to generating isomorphous heavy-atom phasing derivatives

Jinghua Lu and Peter D. Sun

Structural Immunology Section, Lab of Immunogenetics, National Institute of Allergy and Infectious Diseases, National Institutes of Health, 12441 Parklawn Drive, Rockville, MD 20852

Abstract

In attempts to replace the conventional trial-and-error heavy-atom derivative search method with a rational approach, we previously defined heavy metal compound reactivity against peptide ligands. Here, we assembled a composite pH and buffer-dependent peptide reactivity profile for each heavy metal compound to guide rational heavy-atom derivative search. When knowledge of the best-reacting heavy-atom compound is combined with mass spectrometry-assisted derivatization, and with a quick-soak method to optimize phasing, it is likely that the traditional heavy-atom compounds could meet the demand of modern high-throughput X-ray crystallography. As an example, we applied this rational heavy-atom phasing approach to determine a previously unknown mouse serum amyloid A2 crystal structure.

Historic perspective

Protein crystallography is an important tool for modern biology. As the demand for structural information increases, so does the need to improve aspects of protein crystallography. In the post-genomic era, advances in automation of protein crystallization, X-ray data collection and software development have dramatically shortened the turnaround time for structure determinations, thus, increased the throughput of structures[1]. In addition, improvements were made in structural solution methods, especially phasing methods [2]. To date, there are three major methods for phasing: the direct, molecular replacement and heavy-atom replacement methods. While the direct phasing method has only limited success in solving protein structures, it is widely used to solve heavy-atom sub-structures. The molecular replacement method has become more successful owing to the vast number of structures deposited in the protein data bank (PDB). Nevertheless, a significant number of structures still require de novo phasing based on heavy-atom replacements.

The first successful application of isomorphous heavy-atom derivative phasing was by Perutz and his colleagues for solution of the hemoglobin structure [3]. In the early years of protein crystallography, the vast majority of protein structures were determined using heavy-atom derivative phasing. The process of obtaining a suitable heavy-atom adduct for phasing, however, took weeks or even months, and often failed after lengthy screenings. To date, heavy-atom derivative phasing is primarily used to solve macromolecular crystal structures

to which no homologous structures available in PDB. However, the time-consuming, trial-and-error method of conventional heavy-atom compound screening poses significant obstacles to rapid structure determination. For example, conventional heavy-atom searches often result in derivative crystals with reduced diffraction quality and are non-isomorphous to native crystals and, thus, marginalizing their usefulness for phasing. The fact that a conventional heavy-atom compound screening process does not guarantee its success has yielded its applications to an alternative selenomethionine substitution method [4], whereby the certainty of derivatization is guaranteed. The selenomethionine method, however, also has shortcomings. In particular, its phasing is limited to medium-sized proteins, and the selenomethionine labeling remains technically challenging for recombinant proteins produced from other than bacteria expression systems. In comparison, the conventional heavy metals are more electron-rich, thus, can provide stronger isomorphous and anomalous signals and are advantageous for large protein structures. For the conventional heavy-atom derivatization method to meet modern crystallographic needs, it is necessary to minimize its traditional detrimental aspects and maximize its rate of success. Here, we summarize our attempts in developing the conventional heavy-atom derivative phasing method into a rapid and rational approach that uses fewer crystals and provides a better quality of phasing.

Conventional heavy-atom derivatization

Heavy-atom compounds react with protein ligands through either electrostatic or covalent interactions. In general, alkaline metals, halides, lanthanides, and some transition metal ions, such as Fe^{2+} and Zn^{2+} , form electrostatic interactions with opposite charged amino acids, Lys, Arg, His, Asp, Glu, Tyr and N- and C-termini [5]. Mercury, platinum, gold and lead compounds frequently form covalent adducts with secondary amines and other polar groups on Met, His, Cys, Arg, Lys and Tyr side chains, although the negatively charged Pt and Au ions, such as $\text{Pt}(\text{CN})_4^{2-}$ and AuCl_4^- , can also bind electrostatically to positively charged amino acid side chains. Detailed information on protein heavy-atom derivatives and their binding geometries are categorized by Islam et al. in Heavy-atom Data Bank (HDB) <http://www.sbg.bio.ic.ac.uk/had/> [6].

The chemistry of heavy-atom reactions is generally well understood and their activities vary substantially depending on pH, buffer and the presence of certain cationic and anionic groups [5, 7]. Buffer and pH affects the ionization of both heavy-atom compounds and their ligands and, thus, the rate of adduct formation. A systematic evaluation of heavy-atom compound reactivity against reactive peptide ligands in various buffer and pH conditions using mass spectrometry also showed significant variation in heavy metal activities [8]. For example, K_2PtCl_6 is less reactive to methionine at pH below 6, but is quite reactive around neutral pH in MES, sodium cacodylate and HEPES buffers [8]. The same compound is highly reactive toward histidine only in MES buffer. The presence of certain cationic and anionic groups often affect a heavy-atom compound solubility, and thus its reactivity. For example, while mercury acetate is soluble to 200 mM in acetate buffer, adding chloride ions will result in the formation of mercury chloride that is much less soluble, thus reducing the reactive mercury concentration. These variations of heavy-atom reactivity render the conventional derivative screenings of the following characteristics.

1. An empirical trial-and-error process that involves collecting many data sets from heavy atom compound soaked crystals. It requires many crystals, and is time consuming.
2. Many heavy atom compounds either result in severe lattice disruptions or cause non-isomorphism between native and derivative crystals.
3. Anomalous and isomorphous phasing data are often inconsistent due to the lack of isomorphism with native crystals.

Heavy-atom compound reactivity to proteins: A composition of peptide reactivity tables

The time-consuming aspect of the empirical trial-and-error heavy-atom screening method hinders its application for high-throughput structure determination. The key, therefore, is to replace the trial-and-error method with a prediction-based rational approach. Few efforts were made toward predicting the heavy-atom reactions. One such effort was formulated in a Heavy-atom Database System, HATODAS II, which suggests potential heavy-atom reagents for the derivatization based on short sequence motifs [9]. However, heavy-atom compounds are generally reactive against individual chemical groups on the surface of proteins with minimal requirement for sequence and spatial arrangement of amino acids. For example, Pt and Au compounds readily form adducts with Met and His residues regardless of their surrounding sequences. With the assumption that a systematic knowledge of pH and buffer-dependent heavy-atom reactivity could serve to predict the likelihood of a successful derivatization by a given compound, we defined individual heavy-atom compound reactivity profiles using short peptides bearing individual reactive amino acids as substrates [8]. A searchable form of the heavy-atom reactivity profiles can be found at: http://exon.niaid.nih.gov/sis/cgi-bin/heavyatom_reactivity.cgi. The application of these amino acids-based heavy-atom reactivity profiles requires the knowledge of amino acid compositions of a target protein. Since most proteins contain some of the reactive amino acids, and heavy-atom compounds react with various amino acids, it is beneficial to derive a composite heavy-atom reactivity profile that combines a compound's reactivity with individual amino acids. Such composite reactivity profiles can be used as a quick and simple guide to heavy-atom reactivity to proteins. For example, potassium tetrachloroplatinate (K_2PtCl_4) reacts with all three (Met, His, and Cys) peptides in HEPES buffer at pH 7 (table 2–4 of Agniswamy et al, 2008). In comparison, dichloro ethylene diamino platinate $(NH_3)_2(CH_2)_2PtCl_2$ reacts weakly with Met and His and not with Cys under the same condition. Therefore, potassium tetrachloroplatinate is expected to be more reactive to a given protein than dichloro ethylene diamino platinate in Hepes buffer at pH 7. To reflect this different likelihood of reactivity with proteins, we numerically combined the heavy-atom reactivity tables for individual amino acids into a weighted single table and displayed the composite reactivity in a blue to red heat-color scheme of increasing reactivity (Table 1). As many mercury compounds are highly reactive toward much less frequent free cysteine residues but much less reactive toward other amino acids, the mercury-cysteine reactivity is listed separately in Table 2 to avoid a bias toward mercury compounds for majority of proteins that lack free cysteine residues. Among the 43 heavy-atom compounds tested for

over 12 different buffer and pH conditions, only 5% of the heavy-atom and buffer combinations are highly reactive, 20% are moderately reactive, and vast majority of combinations do not yield significant adducts and, thus, are unlikely to produce good phasing derivatives. Given that heavy-atom compounds are generally more reactive to amino acids in small peptides than in proteins due to their restricted accessibilities, the likelihood of finding a highly reactive compound through random screening is expected to be even less. These results emphasize the need for a rational heavy-atom screening process guided by known heavy-atom compound reactivity.

As for individual buffers, our results show that MES at pH 6 is the most reactive condition. For crystals obtained in the MES condition, highly reactive Pt, Au, Hg and Pb compounds are readily available for derivatizations. In contrast, heavy-atom compounds are much less reactive in Tris buffer at pH 8.5 with AuCl_4^- being the most reactive group. Mercury compounds are largely ineffective in the Tris condition. For chicken lysozyme, which is crystallized in sodium acetate at pH 4.8, Table 1 shows approximately a dozen compounds exhibiting better than mild reactivity. These include K_2PtBr_6 , K_2PtCl_4 , K_2PtBr_4 , $\text{Pt}(\text{NH}_3)_2(\text{NO}_2)_2$, $(\text{NH}_4)_2\text{PtCl}_4$, NaAuCl_4 , AuCl_3 , $\text{CH}_3\text{HgCH}_3\text{COO}$, $\text{C}_2\text{H}_5\text{HgPO}_4$, $\text{Pb}(\text{CH}_3\text{COO})_2$ and $\text{Pb}(\text{NO}_3)_2$. Most of these compounds form adducts with lysozyme as revealed by mass spectrometry [8]. However, only three of the platinum compounds, K_2PtBr_6 , K_2PtCl_4 and K_2PtBr_4 , were reported for structure phasing of lysozyme [7], illustrating the ability of these heavy-atom reactivity profiles to generate more useful phasing derivatives. For example, lead acetate derivatized lysozyme crystal but was not reported by earlier random screening [10].

Since peptides are generally more reactive than proteins due to their solvent accessible reactive amino acids, the heavy-atom reactivity profiles represent the most optimal scenario for proteins and can be used to exclude less reactive compounds for derivatization. For example, K_2PtBr_4 is highly reactive to peptides in sodium acetate, but only mild reactive to lysozyme [8]. To further rank the potential derivative compounds among highly reactive chemicals, we routinely use mass spectrometry to screen for derivative adducts in solution [11]. In addition, the current heavy-atom reactivity profiles are derived from heavy-atom compound reactivity to Met, His, Cys, Asp, Glu, Asn, Gln and Tyr residues. While these are often the major heavy-atom binding ligands, other amino acids, such as Lys, Arg, Ser and Thr, are also found to be reactive toward some compounds [5]. A complete profile requires the inclusion of all possible reactive amino acids. Nevertheless, the reactive groups, namely the presence of amines, carboxylates and carbonyls, are represented in the current group of residues. Although the phasing quality of a derivative can only be assessed through X-ray diffractions, the combination of heavy-atom reactivity profiles and mass spectrometry provides a powerful and time saving approach to select the best possible heavy-atom compounds.

Quick-soak optimizes binding, preserves crystal diffraction and isomorphism

An ideal heavy-atom derivative for phasing is the one that has 100% heavy-atom site occupancy, preserves diffraction resolution and remains isomorphous to native crystals.

These are often competing attributes in derivative screenings. In theory, the rate of a chemical reaction is proportional to both reactant concentrations. An optimum reaction rate is achieved when both reactants are in stoichiometric concentrations. Else, the lower concentration component limits the reaction rate. For majority of protein crystals with Matthews' volume, V_m , between 2–4 Å³/Dalton, the crystal lattice protein concentrations are 5–20 mM for typical 20–80 kD proteins. In the past, conventional heavy-atom soakings were often carried out in days and sometime weeks with very limited (<1 mM) concentrations of heavy-atoms [5]. Low heavy-atom compound concentrations help to preserve diffraction but may lead to lower binding occupancies. Conversely, higher compound concentrations favor complete derivatization, but are more disruptive to crystal lattices. Halide soaking is such an example, in which the concentrations of halide salts are 0.5 M or higher in order to saturate the low affinity anionic binding sites in crystals [12]. Under such conditions, a crystal lattice generally deforms in less than a few minutes. The quick-soak method, in which reactions are completed in ~10 minutes using 10 mM heavy-atom compounds, uses a near equal molar heavy-atom:protein concentrations, and thus, achieves optimal reaction kinetics. This also suggests no significant benefit in reaction kinetics with further increased heavy-atom concentrations as the rate limiting factor would be proteins in crystal lattice. Interestingly, we have not seen significant lattice disruptions under quick-soak conditions and all derivatives were isomorphous to their native crystals [10]. In fact, shorter soaking time produced a successful HgCl₂ derivatized FcγRIII crystal, whereas a 4 hour reaction of the same compound resulted in crystal lattice disorder [13, 14]. It is possible that many conventional lattice-disruptive heavy-atom compounds may yield productive derivatives under the quick-soak method.

In general, the quick-soak method minimized the major detrimental effects associated with conventional heavy-atom soaking, such as the deleterious effect on lattice diffraction and the introduction of non-isomorphism. It represents a reaction kinetics optimized, true isomorphous derivatization method.

Heavy-atom derivatization and solution of the mouse SAA2 structure

Previously, we have tested the rational heavy-atom derivative screening method using lysozyme as an example [8, 10]. The result showed that majority of the heavy-atom compounds previously used for lysozyme structure determination can be predicted based on their reactivity profiles and they resulted in good derivatives by the quick-soak method.

Here, we applied this method to de novo heavy-atom phasing and structural solution of a mouse Serum Amyloid A2 (SAA2). Members of Serum amyloid A form a family of inflammatory-linked acute phase proteins and are the underlying cause of secondary amyloidosis. To date, there is no three dimensional structural information available for this family of proteins. We have recently expressed a recombinant mouse SAA2, and obtained well diffracting crystals in PEG 4000, 0.1 M sodium citrate (pH5.6). The crystals belong to a hexagonal space group of P6₄, and diffracted to 2.0-Å resolution. Based on the heavy-atom reactivity profile (Table 1), we selected two highly reactive compounds, K₂PtCl₄ and KAuCl₄, one mildly reactive compound, mercury acetate and one least reactive compound trimethyllead acetate (TMLA) for further assessment of their derivatization of SAA2 in

solution by mass spectrometry [11]. In brief, the heavy-atom reactions were carried out in 10 μ l volume with 5 μ g SAA2 in 20 mM sodium citrate at pH 5.6 mixed with 1–10 mM concentrations of heavy-atom solutions for 10–30 minutes. The formation of SAA2 and heavy-atom adducts was examined by an electrospray ionization mass spectrometer (ESI-MS). SAA2 reacted readily with both K_2PtCl_4 and $KAuCl_4$, forming numerous adducts at 10 mM and 1–5 binding sites at 0.1 mM concentration of the heavy-atom solutions (Figure 1a,b). Mercury acetate showed ~3 binding sites to SAA2 at 10 mM concentration and only one binding site was observed with 50 mM TMLA (Figure 1c,d). All four compounds were used to derivatize SAA2 crystals by the quick-soak method for X-ray data collection. Since all four compounds are more active in acetate than citrate buffer (Table 1), the sodium citrate buffer was replaced with sodium acetate at the same pH in the quick-soak experiments. SAA2 crystals tolerated this buffer change in all quick-soakings. X-ray data were collected at SER-CAT beamlines and processed with HKL2000 [15] (Table 3). The derivative crystals diffracted to the native resolution with no detectable changes in cell dimensions as in other previously observed cases [13, 14]. Among the four heavy-atom compounds, the $KAuCl_4$ derivative resulted in two readily interpretable major binding sites using AutoSharp and good phasing statistics from a combined SIR and MAD phasing [16] (Table 3). The experimental phasing electron density map is of good quality and can be easily traced by Arp/wArp [17] (Figure 2). The structure of SAA2 will be published elsewhere.

Summary

To improve the conventional heavy-atom replacement method suitable for rapid structure determination, we developed a heavy-atom compound reactivity profile-based method to guide derivative screening for proteins. These heavy-atom reactivity profiles, combined with mass spectrometry assessment and quick-soak derivatization, has the potential to transform conventional heavy-atom phasing into a rapid and rational isomorphous phasing method. As an example, we successfully applied this method to a de novo phasing structure of the mouse SAA2.

Acknowledgments

We thank Mr. Mark Garfield of Research Technology Branch of NIAID for performing the mass spectrometry experiments. Ms. Ann Hanson for manuscript editing. All X-ray diffraction data were collected on SER-CAT beamlines in the Advanced Photon Source at Argonne National Laboratory. The use of the Advanced Photon Source was supported by the U. S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract No. W-31-109-Eng-38. The funding for this work was provided by the Division of Intramural Research Program, National Institute of Allergy and Infectious Diseases, National Institutes of Health.

References

1. Dauter Z. Current state and prospects of macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr*. 2006; 62:1–11. [PubMed: 16369088]
2. Beck T, Gruene T, Sheldrick GM. The magic triangle goes MAD: experimental phasing with a bromine derivative. *Acta Crystallogr D Biol Crystallogr*. 2010; 66:374–80. [PubMed: 20382990]
3. Green DW, Ingram VM, Perutz MF. The structure of haemoglobin IV. Sign determination by the isomorphous replacement method. *Proc Roy Soc A*. 1954; 225:287–307.

4. Hendrickson WA, Horton JR, LeMaster DM. Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. *EMBO J.* 1990; 9:1665–72. [PubMed: 2184035]
5. Blundell, TL.; Johnson, LN. Protein crystallography. Academic Press; New York: 1976.
6. Islam SA, Carvin D, Sternberg MJ, Blundell TL. HAD, a data bank of heavy-atom binding sites in protein crystals: a resource for use in multiple isomorphous replacement and anomalous scattering. *Acta Crystallogr D Biol Crystallogr.* 1998; 54:1199–206. [PubMed: 10089497]
7. Blake CC. The preparation of isomorphous derivatives. *Adv Protein Chem.* 1968; 23:59–120. [PubMed: 4882250]
8. Agniswamy J, Joyce MG, Hammer CH, Sun PD. Towards a rational approach for heavy-atom derivative screening in protein crystallography. *Acta Crystallogr D Biol Crystallogr.* 2008; 64:354–67. [PubMed: 18391402]
9. Sugahara M, Asada Y, Ayama H, Ukawa H, Taka H, Kunishima N. Heavy-atom Database System: a tool for the preparation of heavy-atom derivatives of protein crystals based on amino-acid sequence and crystallization conditions. *Acta Crystallogr D Biol Crystallogr.* 2005; 61:1302–5. [PubMed: 16131765]
10. Joyce MG, Radaev S, Sun PD. A rational approach to heavy-atom derivative screening. *Acta Crystallogr D Biol Crystallogr.* 2010; 66:358–65. [PubMed: 20382988]
11. Sun PD, Hammer CH. Mass-spectrometry assisted heavy-atom derivative screening of human Fc gamma RIII crystals. *Acta Crystallogr D Biol Crystallogr.* 2000; 56:161–8. [PubMed: 10666595]
12. Dauter Z, Dauter M, Rajashankar KR. Novel approach to phasing proteins: derivatization by short cryo-soaking with halides. *Acta Crystallogr D Biol Crystallogr.* 2000; 56:232–7. [PubMed: 10666615]
13. Sun PD, Radaev S. Generating isomorphous heavy-atom derivatives by a quick-soak method. Part II: phasing of new structures. *Acta Crystallogr D Biol Crystallogr.* 2002; 58:1099–103. [PubMed: 12077428]
14. Sun PD, Radaev S, Kattah M. Generating isomorphous heavy-atom derivatives by a quick-soak method. Part I: test cases. *Acta Crystallogr D Biol Crystallogr.* 2002; 58:1092–8. [PubMed: 12077427]
15. Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. *Meth Enzymol.* 1997; 276:307–326.
16. Bricogne G, Vonrhein C, Flensburg C, Schiltz M, Paciorek W. Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. *Acta Crystallogr D Biol Crystallogr.* 2003; 59:2023–2030. [PubMed: 14573958]
17. Langer G, Cohen SX, Lamzin VS, Perrakis A. Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat Protoc.* 2008; 3:1171–1179. [PubMed: 18600222]

Figure 1. ESI-MS mass spectra of solution heavy-atom abduct formation between SAA2 and (a) 0.1 mM K_2PtCl_4 , (b) 0.1 mM K_2AuCl_4 , (c) 10 mM Hg acetate and (d) 50 mM TMLA. The molecular weight of native mouse SAA2 is 11932 Daltons.

Figure 2.
A segment of experimental phasing electron density map of SAA2 contoured at 1 σ level.

Table 1

Composite heavy-atom reactivities to M, C, H, D, E, N, Q, Y residues

Heavy-atom compounds	Sodium Acetate		Sodium Citrate			Mes	Sodium Cacodylate			HEPES		Tris-HCl	
	4	5	4	5	6		4	5	6	7	8	7.5	8.5
pH													
Platinum potassium thiocyanate													
Platinum potassium iodide													
Potassium hexabromo platinate													
Potassium tetrachloro platinate													
Potassium tetrabromo platinate													
Diamino platinum dinitrate													
Potassium tetranitro platinate													
Dichloro ethylene diamino platinate													
Potassium tetraacyano platinate													
Ammonium tetrachloro platinate													
Potassium hexachloro platinate													
Potassium tetrachloro aurate													
Sodium tetrachloro aurate													
Gold (III) chloride													
Gold chloride													
Gold potassium bromide													
Potassium tetrabromo aurate													
Gold sodium thiosulfate													
Gold (I) potassium cyanide													
Gold potassium thiocyanide													
Thiomersal													
Mersalyl													
Mercury (II) acetate													
PCMBs													
Potassium mercury tetraiodide													
Methyl mercury (II) chloride

Table 2

Hg compounds reactivities to free cysteine (-SH)

Heavy-atom compounds	Sodium Acetate		Sodium Citrate			MES	Sodium Cacodylate			HEPES		Tris-HCl	
	4	5	4	5	6		5.5	6.5	7	8	7.5	8.5	
pH													
Thiomersal													
Mersalyl													
Mercury (II) acetate													
PCMBs													
Potassium mercury tetraiodide													
Methyl mercury (II) chloride													
Methyl mercury (II) acetate													
Ethyl mercury (II) phosphate													
Mercury (II) cyanide													
Mercury (II) bromide													
Mercury (II) chloride													
Methyl mercury (II) bromide													
p-Chloro mercuric benzoic acid													
Ethyl mercuric chloride													
Mercury (II) iodide													
Mercury oxide													

Table 3

Data collection and phasing statistics of mouse SAA2 crystals

Data collection			
	Native	KAuCl ₄ (Au Peak)	KAuCl ₄ (Au Edge)
Space group	P6 ₄	P6 ₄	P6 ₄
Wavelength (Å)	1.00	1.03926	1.040
Unit cell dimension (Å)	a=b=74.5, c=43.9	a=b=74.1, c=44.1	a=b=74.4, c=44.0
Resolution range (Å) *	30.0–2.00 (2.03–2.00)	50.0–2.60 (2.64–2.60)	50.0–2.60 (2.64–2.60)
Unique reflections	9,512(391)	4,361(209)	4,350(210)
Average redundancy	9.9(2.9)	22.1(22.4)	31.9(32.9)
R _{sym} (%) ^a	11.4(35.3)	11.7(60.2)	11.4(74.7)
I/σ(I)	22.9(3.0)	36.5(5.6)	47.3(7.8)
Completeness (%)	99.3(85.7)	99.9(99.8)	99.9(99.8)
Phasing statistics of KAuCl ₄			
Heavy-atom peak height		Site 1 (15σ)	Site 2 (14σ)
Phasing power ISO[acentric/centric]		2.15/1.54	1.58/1.18
Phasing power ANO		1.18	0.57
Figure-of-merit[acentric/centric]	0.32/0.41		

$$^a R_{\text{sym}} = \sum_h \sum_l |I_l(hkl) - \langle I(hkl) \rangle| / \sum_h \sum_l I_l(hkl),$$

* Values in parentheses represent data for the highest resolution shell.