Environmental Science & Technology

pubs.acs.org/est

Article

Demethylation of Methylmercury in Bird, Fish, and Earthworm

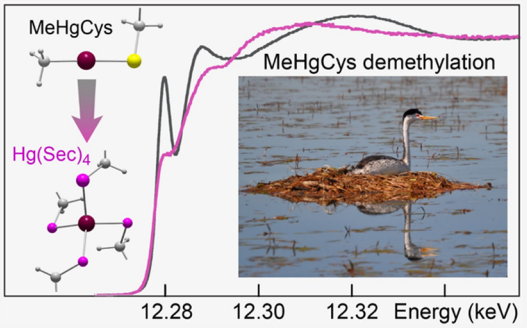
Alain Manceau,* Jean-Paul Bourdineaud, Ricardo B. Oliveira, Sandra L.F. Sarrazin, David P. Krabbenhoft, Collin A. Eagles-Smith, Joshua T. Ackerman, A. Robin Stewart, Christian Ward-Deitrich, M. Estela del Castillo Busto, Heidi Goenaga-Infante, Aude Wack, Marius Retegan, Blanka Detlefs, Pieter Glatzel, Paco Bustamante, Kathryn L. Nagy, and Brett A. Poulin*

Cite This: *Environ. Sci. Technol.* 2021, 55, 1527−1534

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Toxicity of methylmercury (MeHg) to wildlife and humans results from its binding to cysteine residues of proteins, forming MeHg-cysteinate (MeHgCys) complexes that hinder biological functions. MeHgCys complexes can be detoxified *in vivo*, yet how this occurs is unknown. We report that MeHgCys complexes are transformed into selenocysteinate [$Hg(Sec)_4$] complexes in multiple animals from two phyla (a waterbird, freshwater fish, and earthworms) sampled in different geographical areas and contaminated by different Hg sources. In addition, high energy-resolution X-ray absorption spectroscopy (HR-XANES) and chromatography-inductively coupled plasma mass spectrometry of the waterbird liver support the binding of $Hg(Sec)_4$ to selenoprotein P and biomineralization of $Hg(Sec)_4$ to chemically inert nanoparticulate mercury selenide (HgSe). The results provide a foundation for understanding mercury detoxification in higher organisms and suggest that the identified MeHgCys to $Hg(Sec)_4$ demethylation pathway is common in nature.



INTRODUCTION

Mercury (Hg) enters ecosystems as inorganic divalent [Hg(II)] and elemental [Hg(0)] forms through atmospheric deposition at the global scale and through anthropogenic point sources at the local scale. Inorganic Hg(II) is subject to methylation [to methylmercury (MeHg)] by microorganisms[1] and is taken up as MeHg-cysteinate (MeHgCys) in multicellular organisms. MeHgCys has been reported in the skeletal muscle of fish[2,3] and whales[4] as well as in human brain[5] and hair.[6] In marine mammals and seabirds, in which the Hg burden is exceptionally high and selenium (Se) is abundant, MeHgCys is understood to be demethylated in the form of HgSe and Hg(S,Se) granules.[7−12] No other detoxified form of MeHgCys is yet known. Attempts to characterize the HgSe granules by X-ray absorption spectroscopy (XAS) have been deceptive. In crystalline HgSe, the Hg atoms are tetrahedrally coordinated to 4 Se atoms at 2.635 Å and surrounded by 12 Hg neighbors at 4.30 Å. While Hg−Se bond lengths between 2.59 and 2.63 Å have been reported in liver and pituitary tissues from the beluga whale and in the human brain by extended X-ray absorption fine structure (EXAFS) spectroscopy,[5,13] no Hg−Hg pairs have ever been detected in animals and humans. The possibility exists that MeHgCys is detoxified to organic selenide forms prior to nanoparticulate HgSe nucleation.

To address this question, we investigated *in vivo* chemical forms of Hg in the liver and extrahepatic tissues of the waterbird Clark's grebe (*Aechmophorus clarkii*) and several species of freshwater fish and in whole bodies of earthworms. The Clark's grebe was collected from Lake Berryessa, which is impacted by mining activity in the California Coast Range (Table 1). Five piscivore fish, including three peacock bass (two *Cichla temensis* and one *Cichla monoculus*) and two pescada branca (*Plagioscion squamosissimus*), were collected in the Tapajós River in Brazil contaminated by artisanal gold mining activity (Tables S1 and S2). Earthworms (n = 13) were collected from two French soils impacted by Hg deposition from a chlor-alkali plant[14] and analyzed with and without digestive tracts (Table S3).

Mercury speciation analysis was performed using high energy-resolution (HR)-XAS.[15] In the HR X-ray absorption near-edge structure (XANES) mode, Hg ligands are identifiable with enhanced chemical sensitivity compared to standard XANES, and Hg−Hg pairs are detectable in biological tissues at sub-ppm Hg concentrations (<1 mg total Hg kg^{-1}),[6,16,17] even when the pairs occur in nanometric HgSe precipitates.[12] HR-XAS also provides high-quality HR-EXAFS data for

Received: July 24, 2020
Revised: December 30, 2020
Accepted: December 30, 2020
Published: January 21, 2021

Table 1. Total Concentrations (mg kg^{-1}) of Hg and Se Determined by Chemical Analysis, and Concentrations of MeHg, Hg Tetraselenolate, and Hg Dithiolate Determined by HR-XANES in Clark's Grebe (mg Hg kg^{-1} Dry Weight)

tissue	$[Hg]_{Tot}$	$[Se]_{Tot}$	[MeHgCys]	$[Hg(Sec)_4]^a$	$[Hg(SR)_2]$
breast feather	40.7 ± 4.1	1.04 ± 0.10	40.7 ± 4.6	0.0 ± 2.1	0.0 ± 2.1
brain	3.1 ± 0.3	1.55 ± 0.16	3.1 ± 0.3	0.0 ± 0.2	0.0 ± 0.2
muscle	6.5 ± 0.7	2.31 ± 0.23	4.3 ± 0.5	0.7 ± 0.3	1.5 ± 0.4
kidneys	21.6 ± 2.2	10.6 ± 1.0	6.0 ± 1.2	12.7 ± 1.7	2.6 ± 1.1
liver	43.1 ± 4.3	19.3 ± 1.9	6.0 ± 2.2	37.1 ± 4.3	0.0 ± 2.1

[a]Minor $Hg_x(Se,Sec)_y/HgSe_{NP}$ species are not modeled independent of $Hg(Sec)_4$ because the amount cannot be quantified.

complementary insight into the bonding environment and local structure of Hg. The parasitic Se K-absorption edge, which limits the energy range of the Hg L_3-edge EXAFS signal to 374 eV ($k \sim 9.6$ Å^{-1})[18−21] unless Ga K-edge absorption filters are used,[22,23] is not observed at high energy-resolution.[24−26] The speciation of Hg and Se in Clark's grebe liver and muscle was also investigated by double affinity chromatography high-performance liquid chromatography coupled to inductively coupled mass spectrometry (AF-HPLC-ICPMS).[27,28]

MATERIALS AND METHODS

Sampling and Sample Preparation. *Clark's Grebe.* An adult male Clark's grebe (*A. clarkii*; band # 2097-56064) was collected from Lake Berryessa (California, United States; 38°36′31″N 122°14′33″W NAD 83) in September 2012[29] and necropsied at the U.S. Geological Survey (Dixon, CA) to obtain the following tissues: head feather, breast feather, brain, pectoral muscle, kidneys, and liver. Tissues were lyophilized and homogenized. Clark's grebes are piscivorous, nest inland on lakes and migrate to coastal climates over the winter, and exhibit some of the highest Hg concentrations among bird species in western North America.[30]

Fish. Peacock bass (*Cichla* sp.), called "tucunaré" in the Amazon region, is a piscivore fish in the adult state, and lakes are its main habitat. The fish species are given in Table S1. Tucunaré weighs up to 10 kg, and there is no protection from harvesting during the period of reproduction, therefore it is sold on local markets all year long. Tucanaré is not a migratory species, and the fish live generally in groups; therefore, the Hg content reflects local dietary intake. *P. squamosissimus*, called "pescada branca" in the Amazon region, is also a piscivore fish which can exceed 10 kg. It is a migratory fish that can swim many kilometers up and down river. The specimens used in this study were fished by artisanal fishermen in September 2018 in lakes near the community of São Luis do Tapajós (4°25′28.8″S 56°14′36.5″W NAD 83) located on the Tapajós River, which borders artisanal gold mining areas where liquid Hg^0 is used for gold mining.

Earthworm. Earthworms were collected in 2017−2018 from two soils (S1: 45°05′13.1″N, 5°44′27.5″E NAD 83; S2: 45°05′26.1″N, 5°44′25.5″E NAD 83) near the chlor-alkali plant in Champ-sur-Drac located ∼10 km south of Grenoble, France. Site S1 is situated approximately 250 m to the southeast and site S2 approximately 300 m to the northeast of the electrolytic cells. Soil samples of the topsoil (A-horizon) were taken for total Hg and Se analyses and pH determination. They were freeze-dried; the undecomposed organic material was removed under a microscope and then sieved to <2 mm and <63 μm. Soil pH was measured on the <2 mm fraction after 45 min agitation of 6 g soil in 15 mL of Milli-Q water. Soil S1 had a pH of 7.7 and soil S2 had a pH of 8.0. Hg and Se concentrations were measured on the <63 μm fraction. Total Hg concentrations of soils were S1 = 0.20 ± 0.01 mg kg^{-1} and S2 = 27.6 ± 0.3 mg kg^{-1}. A concentration of 0.23 ± 0.02 mg kg^{-1} was reported previously in a tilled soil layer situated 900 m north of the plant.[14] The origin of Hg at this site is probably atmospheric. The high concentration at the second site probably results from a Hg spill. The Hg concentration was high enough for the determination of Hg speciation by HR-XANES spectroscopy. Figure S1 shows that Hg is mainly in the form of nanoparticulate β-HgS. Total Se concentration of soil was lower at site S1 (0.36 ± 0.04 mg kg^{-1}) compared to that at S2 (0.78 ± 0.08 mg kg^{-1}).

Five endogeic and one anecic earthworms were sampled at site S1 and five endogeic and two anecic at site S2 (Table S3). Eleven out of the 13 earthworms were rapidly frozen in liquid nitrogen, freeze-dried, and crushed in a mortar for chemical and HR-XANES analyses. The digestive tracts were removed from one endogeic earthworm from site S1 (EarthW-endo-1, 1.5 ± 0.2 mg Hg kg^{-1} dw) and one endogeic (EarthW-endo-9, 5.4 ± 0.5 mg Hg kg^{-1} dw) and one anecic (EarthW-anecic-1, 1.2 ± 0.1 mg Hg kg^{-1} dw) earthworm from site S2, and the dissected earthworms were freeze-dried. All dry powders were preserved in a desiccator until analysis. The total Hg concentrations of the 13 earthworms ranged from 0.53 to 10.1 mg kg^{-1}. The mean Hg concentration of endogeic earthworms is 1.9 ± 1.0 mg kg^{-1} ($n = 5$) at site S1 and 5.8 ± 2.5 mg kg^{-1} ($n = 5$) at site S2. Earthworms from the S2 site contained higher proportion of MeHg than those from the S1 site.

Methods. A detailed description of the experimental methods, including chemical, polymerase chain reaction (PCR), and protein analyses, SelP amino acid sequences, and HR-XANES and HR-EXAFS measurements and data reduction, is given in the Supporting Information. Briefly, all samples were measured for total Hg and Se concentrations by chemical analysis and for Hg speciation by HR-XANES spectroscopy. Only the Clark's grebe liver tissue contained high enough Hg (43.1 ± 4.3 mg kg^{-1}) for HR-EXAFS measurement. The speciation of Hg and Se in Clark's grebe liver and muscle tissues was also measured by AF-HPLC-ICPMS. PCR analysis was performed on soil and earthworm microflora extracted from the digestive tract for the presence of Hg-methylation genes. The predicted tertiary structures of SelP amino acid sequences across fish, birds, and mammals were modeled. Data generated during this study are available at https://doi.org/10.5066/P96NP376.[31]

RESULTS

Mercury and Selenium Concentrations. The Clark's grebe tissues exhibited total Hg and Se concentrations consistent with grebes from high Hg locations in the western United States,[32] and the hierarchy in both element

https://dx.doi.org/10.1021/acs.est.0c04948
Environ. Sci. Technol. 2021, 55, 1527−1534

concentrations was similar to those of other vertebrates (brain < muscle < kidneys < liver; Table 1).[33] The Clark's grebe feathers had elevated total Hg (40.7 ± 4.1 mg kg^{-1}) and low Se levels (1.04 ± 0.10 mg kg^{-1}). In the three fish species, the total Hg and Se concentrations were comparable (mean [Hg] = 2.6 ± 1.4 mg kg^{-1}, mean [Se] = 3.0 ± 3.0 mg kg^{-1}), Se concentrations were higher in liver compared to muscle (mean $[Se]_{liver}$ = 4.9 ± 3.3 mg kg^{-1}, mean $[Se]_{muscle}$ = 1.0 ± 0.1 mg kg^{-1}), and the relative concentration of total Hg in liver versus muscle samples varied across the five individuals (from 0.4 to 1.8). The total Hg and Se concentrations in earthworms ranged from 1.2 to 10.1 and 1.0 to 28.3 mg kg^{-1}, respectively (mean [Hg] = 3.7 ± 2.5 mg kg^{-1}, mean [Se] = 10.6 ± 8.0 mg kg^{-1}) (Tables S2 and S3).

Identification of the Hg-Selenocysteinate Complex in Clark's Grebe. Two dominant Hg species occur in the Clark's grebe tissues as indicated by isosbestic points in the HR-XANES spectra, points where all spectra have the same normalized absorption (Figure 1a). The feathers and brain spectra have an intense near-edge peak at 12,279.8 eV characteristic of MeHgCys[6,34] (Figure S2a). Conversely, the liver spectrum has a weak near-edge peak, and the top edge occurs at lower energy (12,312−12,338 eV). These differences are diagnostic of a shift in Hg coordination from linear in feathers and brain to tetrahedral in liver (Figure S2b).[35] Muscle and kidney tissues contain variable proportions of these two species. The liver spectrum does not match exactly any known Hg spectrum and, of the tetrahedral compounds in our extended spectral database (ref 35 and refs therein), corresponds most closely to nanoparticulate tiemannite ($HgSe_{NP}$) (Figure 1b). However, the spectrum lacks oscillations in the top-edge region (12,300−12,340 eV) that arise from multiple scattering events of the photoelectron on the 12 Hg atoms surrounding a central Hg atom in HgSe. Such oscillations are intense in well-crystallized HgSe and less intense in nanoparticulate materials, which have wide ranges of angles and bond lengths as commonly observed for metal clusters in proteins.[35] Instead, the top-edge region is bell-shaped, which indicates that little of the Hg occurs as HgSe granules of any size. The seleniol nature of the new tetrahedral Hg species was confirmed by HR-EXAFS spectroscopy (Figure 2). The liver HR-EXAFS data were fit with Hg bound on average to 2.6 ± 0.6 Se atoms at 2.61 ± 0.01 Å and 0.8 ± 0.2 S atoms at 2.35 ± 0.03 Å in the first coordination shell,

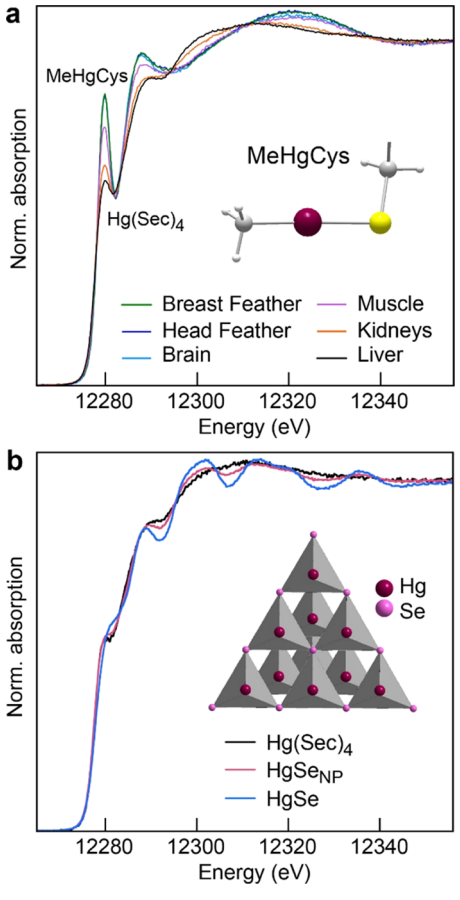


Figure 1. Mercury coordination in the Clark's grebe derived from Hg L_3-edge HR-XANES. (a) Spectra from the grebe tissues. (b) Spectrum of the Hg tetraselenocysteine complex [$Hg(Sec)_4$] with spectra from well-crystallized and nanoparticulate HgSe ($HgSe_{NP}$). The linear coordination of Hg in MeHgCys and a portion of the HgSe structure are represented in inset. Dark red, purple, yellow, gray, and light gray spheres represent Hg, Se, S, C, and H, respectively.

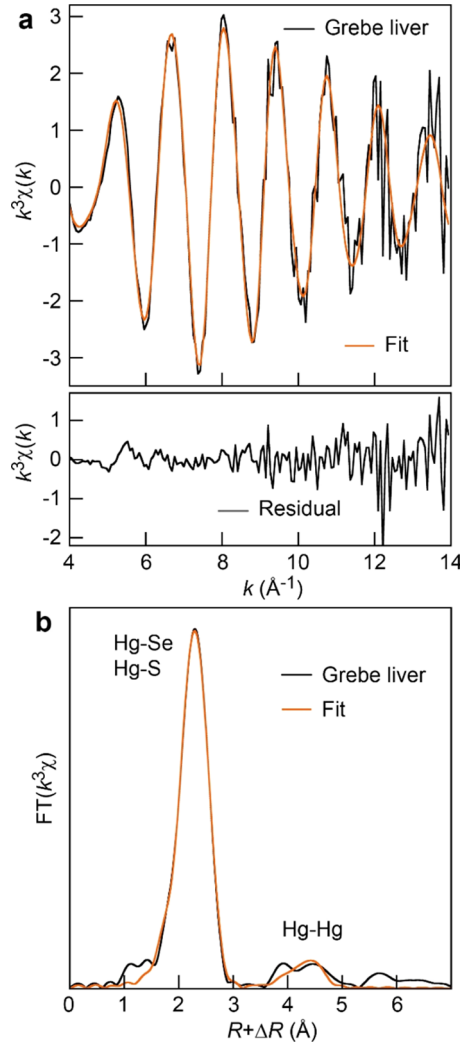


Figure 2. Mercury L_3-edge HR-EXAFS data of the Clark's grebe liver. (a) HR-EXAFS spectrum fit with Hg−Se and Hg−S pairs in the first coordination shell and Hg−Hg-pairs in the second coordination shell, and fit residual. (b) Fourier transform of the HR-EXAFS spectrum and fit. Interatomic distances are not corrected for phase-shift (ΔR).

https://dx.doi.org/10.1021/acs.est.0c04948
Environ. Sci. Technol. 2021, 55, 1527−1534

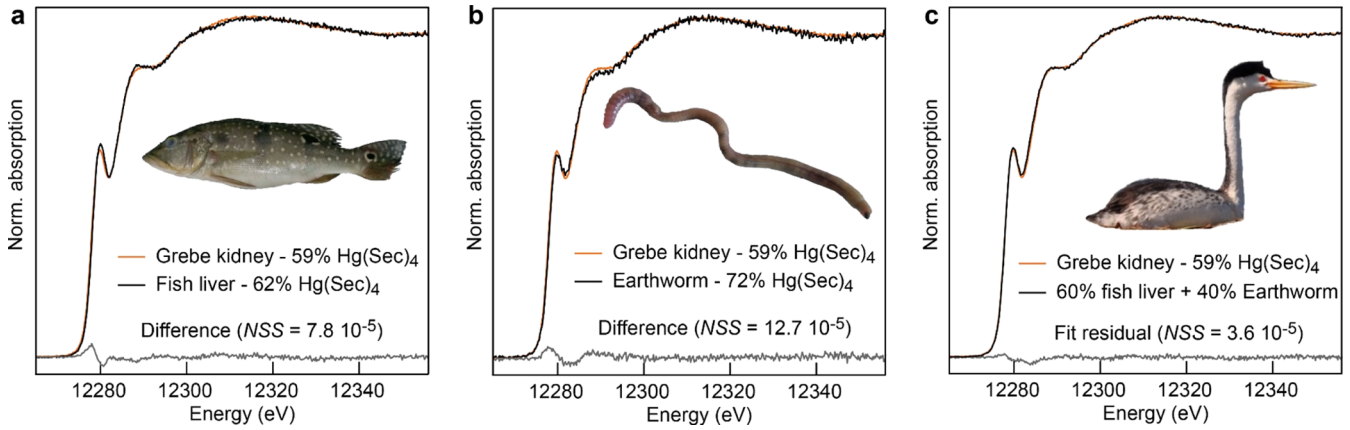


Figure 3. Mercury coordination in peacock bass and earthworm derived from Hg L_3-edge HR-XANES. (a) Comparison of the fish liver spectrum from the individual having the highest proportion of $Hg(Sec)_4$ (PeacockBass-3-L; 62 ± 5%, Table S2) with the spectrum from the Clark's grebe kidneys [59 ± 5% $Hg(Sec)_4$]. (b) Comparison of the earthworm spectrum from the individual having the highest proportion of $Hg(Sec)_4$ (72 ± 10%, Table S3) with the spectrum from the Clark's grebe kidneys. (c) Fit of the Clark's grebe kidneys spectrum with 60% fish spectrum and 40% earthworm spectrum. The calculated proportion of $Hg(Sec)_4$ is (0.60 × 0.61) + (0.40 × 0.69) = 64%, close to the nominal value of 59%. *NSS* is the normalized sum-squared residual expressed as $\Sigma[(y_2 - y_1)^2]/\Sigma(y_2^2)$.

consistent with a four-fold coordination of Hg to Se[36] for the dominant seleniol species (Figure S3a) and a two-fold coordination of Hg to S for minor MeHgCys (Table S4). The minor contribution of MeHgCys was confirmed at 18% by chemical analysis (Figure S4). We conclude that the inorganic Hg species is a tetrahedral selenolate complex with selenocysteine residues [$Hg(Sec)_4$].

The HR-XANES spectrum of $Hg(Sec)_4$ was obtained by the iterative transformation factor analysis of all data from bird, fish, and earthworm tissues, in which varying proportions of the $Hg(Sec)_4$ complex were observed together with MeHgCys and a Hg dithiolate complex $Hg(SR)_2$ (Figure S5). Cysteine is the most likely biological thiol associated with either a peptide (*e.g.*, glutathione) or protein. Using the $Hg(Sec)_4$ spectrum, the best two-component fit of the grebe liver spectrum was obtained with 86 ± 5% $Hg(Sec)_4$ + 14 ± 5% MeHgCys, the latter represented by the Clark's grebe feather (Figure S6a and Table S5). Although the $Hg(Sec)_4$ spectrum, like the grebe liver spectrum, lacks the oscillations in the top-edge region observed in $HgSe_{NP}$ (Figure 1b), several lines of evidence point to the copresence of individual $Hg(Sec)_4$ complexes and disordered $Hg_x(Se,Sec)_y$ clusters in the grebe liver, as in hepatic and extrahepatic tissues of giant petrels *Macronectes* spp.[12] Careful study of the top-edge region of the $Hg(Sec)_4$ spectrum shows that it slightly deviates from the Clark's grebe liver spectrum at energies where HgSe exhibits intense oscillations (arrows in Figure S6a). For this reason, a modest improvement in fit quality (10%) was achieved by including 6% HgSe in the fit of the grebe liver (Figure S6b), which decreased the proportion of $Hg(Sec)_4$ from 86 to 80%. It is likely that more than 6% of the total Hg is present as $Hg_x(Se,Sec)_y$ clusters in the grebe liver as (1) the HgSe reference is a proxy for well-crystallized HgSe granules and does not account for disordered $Hg_x(Se,Sec)_y$ clusters (Figure S7) and (2) the molar ratio of Se and Hg (Se/Hg_{tot}) in the grebe liver is 1.1 ± 0.2 (Figure S8a). The existence of $Hg_x(Se,Sec)_y$ clusters in the grebe liver is further supported by the second coordination shell fitting of the HR-EXAFS spectrum, which detects 2.1 Hg−Hg pairs at 4.46 ± 0.03 Å compared to 12 pairs at 4.30 Å in crystalline HgSe (Figures 2b, S9 and Table S4).

Mercury Speciation in the Two Animal Phyla. The $Hg(Sec)_4$ complex was observed in the two animal phyla, and proportions of Hg species were obtained by least-squares fitting of all HR-XANES spectra with the linear combinations of MeHgCys, $Hg(Sec)_4$, and $Hg(SR)_2$. In the Clark's grebe, the amounts of mercury in the major $Hg(Sec)_4$ complex plus minor $Hg_x(Se,Sec)_y/HgSe_{NP}$ are 37.1 ± 4.3 mg kg^{-1} in liver, 12.7 ± 1.7 mg kg^{-1} in kidneys, and 0.7 ± 0.3 mg kg^{-1} in muscle (Table 1 and Table S5). The $Hg(Sec)_4$ species is also present in fish liver and earthworm samples in varying mixtures with MeHgCys and $Hg(SR)_2$ (Tables S2 and S3), but fish muscle tissues contain exclusively MeHgCys, in line with previous XANES studies.[2,3] In the fish, the proportion of $Hg(Sec)_4$ in the liver correlates positively with the $[Hg]_{liver}$:$[Hg]_{muscle}$ ratio (Figure S10) and is as high as 62 ± 5% in the liver of the fish with the highest Hg concentration (*C. temensis*, 5.5 mg Hg kg^{-1}), suggesting that $Hg(Sec)_4$ is a product of demethylation. The MeHgCys and $Hg(Sec)_4$ species occur in all earthworms in proportions of 17−60 ± 10% and 10−72 ± 10%, respectively, whereas $Hg(SR)_2$ is either absent or present in proportions up to 73 ± 10% (Table S3). Total inorganic Hg [$Hg(Sec)_4$ and $Hg(SR)_2$] in earthworms ranges from 39 to 83%, which is lower than the 90−94% reported in earthworms from 34 natural forest soils throughout Switzerland,[37] suggesting that Hg-methylation occurs in the two contaminated soils. The *hgcA* gene, which is required for Hg-methylation, was observed in the soil and within the digestive tracts of the earthworms (Figures S11 and S12 and Table S6). The majority of MeHg in earthworm samples is interpreted to be in the tissue because earthworms with and without digestive tracts had comparable amounts of MeHgCys (Table S3).

Spectra with similar proportions of $Hg(Sec)_4$ from the two phyla show a high degree of similarity, as demonstrated for the fish liver (62 ± 5%, PeacockBass-3-L), an earthworm (72 ± 10%, EarthW-anecic-2), and Clark's grebe kidneys (59 ± 5%) (Figure 3a,b). The proportion of $Hg(Sec)_4$ derived from a linear least-squares fit of the fish and earthworm spectra to the grebe kidneys spectrum is 64% (Figure 3c). Agreement between the separate estimates of 64% and 59% $Hg(Sec)_4$ in the Clark's grebe kidneys serves as an internal consistency check on the quantification of the Hg species (Tables 1 and

https://dx.doi.org/10.1021/acs.est.0c04948
Environ. Sci. Technol. 2021, 55, 1527−1534

S5). Further, the accuracy of the spectral fit results was evaluated by comparing the fractional amount and concentration of MeHg obtained by HR-XANES versus chemical extraction and atomic fluorescence detection (Figure S4 and Table S7). Good agreement was observed between the two independent methods for relative abundance and concentration of MeHg (Figure S4).

DISCUSSION

The data provide strong evidence for demethylation of MeHg as a tetraselenolate complex in the two animal phyla. This species can be formed with low-molecular-weight (LMW) seleniol molecules and with selenoproteins. Selenoneine is a LMW Se-containing molecule[38] small enough to form a $Hg(Sec)_4$ complex (Figure S3b), as demonstrated with a synthetic analogue.[36] Of selenoproteins, only selenoprotein P (SelP) contains at least four selenocysteine residues.[39] Selenoneine has been reported in fish,[38] and Hg-bound SelP was identified in the plasma of Inuit adults[40] and miners exposed to high concentrations of Hg.[41]

SelP is mainly produced in the liver and secreted into the blood plasma where it transports Se to peripheral tissues.[33,42] SelP expression was highly upregulated in the miners, supporting the idea that SelP may protect against Hg toxicity. SelP is also produced in lesser amounts in extrahepatic tissues. In vertebrates (goat—*Capra hircus*, mouse—*Mus musculus*), mRNA expression of SelP decreased in the order liver > kidneys > muscle > brain.[43,44] The variation of the biosynthesis activity of SelP mRNA among tissues coincides with the percentage fraction of $Hg(Sec)_4$ in the Clark's grebe tissues derived from HR-XANES: 86 ± 5% in liver, 59 ± 5% in kidneys, 11 ± 5% in muscle, and ≤5% in brain (Table S5). The speciation analysis of protein extracts[27] of Clark's grebe liver and muscle tissues by AF-HPLC-ICPMS supports association of Hg and Se with SelP (Figure 4). Good agreement between biochemistry and spectroscopy is strong evidence for the binding of Hg to selenoprotein P in the cells of the Clark's grebe and fish. Similar binding configurations have been identified between Sec residues and metal coenzymes.[45] Less conclusive is the selenolate ligand in the earthworm, given the absence of SelP in annelids;[46] perhaps, it is selenoneine.

Multiple selenocysteine (Sec) alignments of SelP from fish, birds, and mammals (including humans) show that the C-terminal domain contains the consensus $XUXUX^6UXUX$ motif (single-letter amino acid code, where U is Sec and X is any other amino acid) (Figure S13). This U- or Sec-rich region is the most plausible binding site for Hg, given the proximity of the protein chain to the cytosol. This hypothesis was evaluated by modeling the tertiary structure of the $XUXUX^6UXUX$ amino acid region of SelP for the great crested grebe (*Podiceps cristatus*) and zebrafish (*Danio rerio*) by iterative threading assembly refinement (I-TASSER,[47] Figures S14 and S15). The three bird and fish models with the highest prediction score all consist of a loop with the X^6 sequence forming the turn and the UXU motifs facing each other on each strand (Figures 5 and S15). The predicted conformations are highly favorable to bind the Hg^{2+} ion in a tetrahedral configuration. The geometric similarity of bird and fish selenolate sites, despite having relatively low amino acid sequence identity (35% in the C-terminal domain), suggests that this binding site is preserved in vertebrates and may have a common origin.

The most direct interpretation of the findings is that the $XUXUX^6UXUX$ SelP site in birds and fish, and selenoneine or

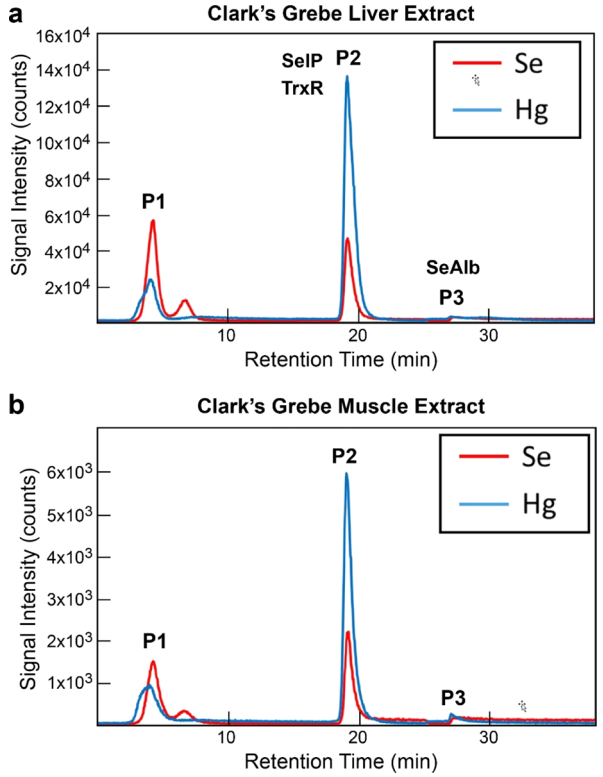


Figure 4. Double affinity chromatograms of Se (red lines) and Hg (blue lines) of Clark's grebe liver extract (a) and Clark's grebe muscle extract (b) by AF-HPLC-ICPMS. The Se species of each peak are as follows: peak 1, P1 (glutathione peroxidase 3; GPX3 and other nonretained species, *e.g.*, inorganic Se and LMW Se species), peak 2, P2 (heparin-binding selenoproteins; likely selenoprotein P, SelP), and peak 3, P3 (Se-albumin; SeAlb). Chromatograms show that, of the Se species, the majority of Hg is associated with P2 (87.8 ± 0.4 and 84.3 ± 0.7% in the liver and muscle, respectively).

an unknown LMW species in earthworms, bind Hg(II) as the product of MeHgCys demethylation. The coexistence of $Hg(SR)_2$ with $Hg(Sec)_4$ in numerous earthworms (Table S3) and in the Clark's grebe muscle and kidney tissues (Table 1) could arise from an additional, though less prominent, demethylation pathway producing $Hg(SR)_2$ or a deficiency in available selenolate. Although the $Hg(SR)_2 \rightarrow Hg(Sec)_4$ reaction cannot be dismissed in theory, the formation of $Hg(Sec)_4$ through this pathway has not been documented. More work is ultimately required to document thiolate versus selenolate competition for MeHg and Hg(II) at the protein level.[23] In terms of reactivity, selenolate (as selenocysteine) outcompetes thiolate (as cysteine) for association with MeHg,[42] and selenolate addition to MeHgCys is exergonic and cleaves both the Hg−C and Hg-SR bonds.[36,48,49] In addition, the protolytic cleavage of the Hg−C bond is promoted when the coordination of Hg increases from 2 in the reactant (MeHgCys) to 4 in the product [$Hg(SR)_4$,[50] and here, $Hg(Sec)_4$]. The strictly conserved $XUXUX^6UXUX$ motif of SelP may serve as a catalytic site for the demethylation of MeHg, analogous to cysteine residues in the organomercurial lyase MerB.[51] MeHg demethylation by selenoneine has been demonstrated *in vitro* in fish blood cells and human embryonic kidney cells.[52]

Placed in a broader context, the results presented here fuel the idea that the C-terminal domain of SelP acts as a nucleation center for the formation of HgSe in vertebrates,

https://dx.doi.org/10.1021/acs.est.0c04948
Environ. Sci. Technol. 2021, 55, 1527−1534

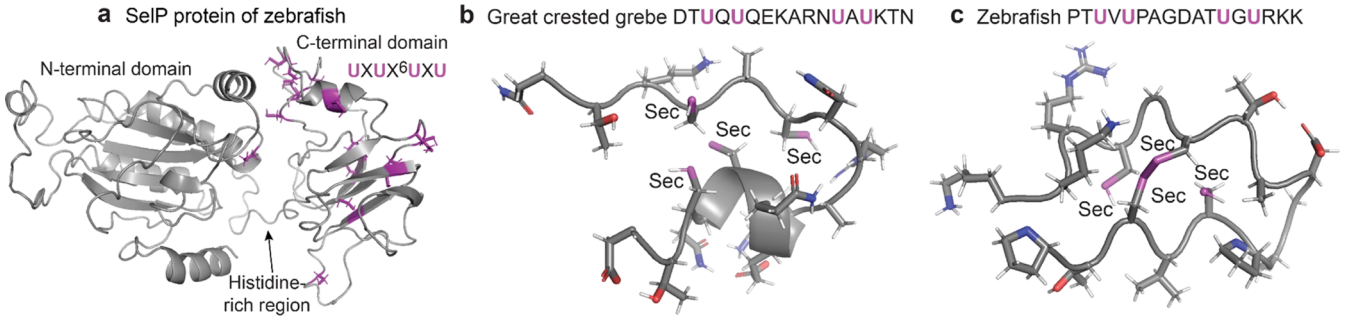


Figure 5. Modeled tertiary structure of SelP. (a) Ribbon representation of the SelP protein for zebrafish (*D. rerio*) as a proxy for peacock bass. (b,c) Amino acid sequence and predicted structure of the UXUX^{6}UXU region of the C-terminal domain of SelP for the great crested grebe (*P. cristatus*) as a proxy for the Clark's grebe (b) and zebrafish as a proxy for peacock bass (*Cichla* sp.) and pescada branca (*P. squamosissimus*) (c). U is the single-letter selenocysteine (Sec) code. Purple, red, and blue represent selenol, carboxyl/hydroxyl, and amino groups from amino acid side chains, respectively. Other models are shown in Figure S16.

driving *in vivo* MeHgCys detoxification. This finding is based on the following observations: (1) Hg is four-coordinate in $Hg(Sec)_4$ and HgSe; (2) the Se/Hg_{tot} molar ratio of the HgSe grains in the liver of pilot whales is above 1 as the grain size diminishes and in the tissue surrounding the grain aggregates,[11] which suggests that the grains are embedded in a SelP-rich matrix having a Se/Hg_{tot} ratio approaching and above 4; (3) here, the $Se/Hg(Sec)_4$ ratio is always greater than 4, as expected if there is a $Hg(Sec)_4$ complex, except in the Clark's grebe liver ($Se/Hg(Sec)_4 = 1.3 \pm 0.2$) and kidneys [$Se/Hg(Sec)_4 = 2.1 \pm 0.3$] (Figure S8b), the former confirmed to contain $Hg_x(Se,Sec)_y/HgSe_{NP}$ (Figure S9); (4) the C-terminal domain of SelP contains six UXU motifs, which could seed the formation of $Hg_x(Se,Sec)_y$ clusters.[12] These considerations show how Hg speciation results, elemental Se/Hg ratios, and protein analyses can be reconciled to provide a foundation for understanding the process of HgSe formation. A similar nucleation mechanism has been proposed for the formation of $Hg_4(Cys)_{12}$ clusters from a $Hg(Cys)_2$ complex and five proximal CXC motifs in metallothionein.[35] In SelP, we propose that Hg binds to the XUXUX^{6}UXUX site at low concentration, forms highly disordered $Hg_x(Se,Sec)_y$ clusters through metallophilic interactions[16,35,53,54] in the C-terminal domain at higher concentration, and finally precipitates as HgSe by self-assembly of the $Hg_x(Se,Sec)_y$ scaffolds in highly exposed or older vertebrates. $Hg_x(Se,Sec)_y$ growth may occur through catalysis of MeHgCys demethylation at nanoparticle surfaces. Structural sketches of the $Hg_x(Se,Sec)_y$ biotemplates, inspired from the topology of the metal clusters in vertebrate, invertebrate, and bacterial metallothioneins, are shown in Figures S7a and S16. Biomineralization by protein assemblies with high symmetry is common in nature.[55] The proposed reaction pathway is probably how Hg is detoxified as HgSe in human brain[5] and seabirds.[12] The binding of Hg to SelP may, however, deplete neuronal and glial cells in organo-Se essential to selenoenzyme synthesis and activity.[42] It may also alter the antioxidant function of SelP.[33,42]

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.est.0c04948.

Materials and methods, supporting tables and figures, and Cartesian coordinates of the $Hg(selenoneine)_4$ complex and the $Hg_{10}(SeMe)_{20}$ cluster (PDF)
HR-XANES spectra (XLSX)

■ AUTHOR INFORMATION

Corresponding Authors

Alain Manceau − *Université Grenoble Alpes, ISTerre, CNRS, Grenoble 38000, France;* orcid.org/0000-0003-0845-611X; Email: alain.manceau@univ-grenoble-alpes.fr

Brett A. Poulin − *U.S. Geological Survey, Boulder 80303, Colorado, United States; Department of Environmental Toxicology, University of California Davis, Davis 95616, California, United States;* orcid.org/0000-0002-5555-7733; Email: bapoulin@ucdavis.edu

Authors

Jean-Paul Bourdineaud − *Université de Bordeaux, Institut Européen de Chimie et Biologie, CNRS, Pessac 33600, France*

Ricardo B. Oliveira − *Universidade Federal do Oeste Pará, LabBBEx, Santarém 68180-000, Pará, Brazil*

Sandra L.F. Sarrazin − *Universidade Federal do Oeste Pará, LabBBEx, Santarém 68180-000, Pará, Brazil*

David P. Krabbenhoft − *Upper Midwest Water Science Center, U.S. Geological Survey, Middleton 53562, Wisconsin, United States;* orcid.org/0000-0003-1964-5020

Collin A. Eagles-Smith − *Forest and Rangeland Ecosystem Science Center, U.S. Geological Survey, Corvallis 97330, Oregon, United States;* orcid.org/0000-0003-1329-5285

Joshua T. Ackerman − *Western Ecological Research Center, U.S. Geological Survey, Dixon 95620, California, United States;* orcid.org/0000-0002-3074-8322

A. Robin Stewart − *U.S. Geological Survey, Menlo Park 94025, California, United States*

Christian Ward-Deitrich − *National Measurement Institute, LGC Limited, Teddington TW11 0LY, Middlesex, U.K.*

M. Estela del Castillo Busto − *National Measurement Institute, LGC Limited, Teddington TW11 0LY, Middlesex, U.K.;* orcid.org/0000-0001-6595-5236

Heidi Goenaga-Infante − *National Measurement Institute, LGC Limited, Teddington TW11 0LY, Middlesex, U.K.*

Aude Wack − *Université Grenoble Alpes, ISTerre, CNRS, Grenoble 38000, France*

Marius Retegan − *European Synchrotron Radiation Facility (ESRF), Grenoble 38000, France;* orcid.org/0000-0002-1525-1094

https://dx.doi.org/10.1021/acs.est.0c04948
Environ. Sci. Technol. 2021, 55, 1527−1534

Blanka Detlefs − *European Synchrotron Radiation Facility (ESRF), Grenoble 38000, France*
Pieter Glatzel − *European Synchrotron Radiation Facility (ESRF), Grenoble 38000, France;* orcid.org/0000-0001-6532-8144
Paco Bustamante − *Université La Rochelle, CNRS, Littoral Environnement et Sociétés, La Rochelle 17000, France;* orcid.org/0000-0003-3877-9390
Kathryn L. Nagy − *Department of Earth and Environmental Sciences, University of Illinois at Chicago, Chicago 60607, Illinois, United States;* orcid.org/0000-0002-4997-7547

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.est.0c04948

Notes

The authors declare no competing financial interest.
All data supporting the findings of this study are available within the paper and have been deposited in the U.S. Geological Survey repository ScienceBase.[31] The deposit includes all HR-XANES spectra, the Hg L_3-edge HR-EXAFS spectrum of the Clark's grebe liver, and the Cartesian coordinates of the $Hg(selenoneine)_4$ complex and the $Hg_{10}(SeMe)_{20}$ cluster.

■ ACKNOWLEDGMENTS

Support was provided to A.M., J.P.B., and K.L.N. by the French National Research Agency (ANR) under grant ANR-12-BS06-0008-01, to A.M. and P.G. by the ANR under grant ANR-10-EQPX-27-01 (EcoX Equipex), to B.A.P. and K.L.N. by the U.S. National Science Foundation under grants EAR-1629698 and EAR-1628956, to B.A.P., D.P.K., A.R.S., C.A.E., and J.T.A. by the U.S. Geological Survey (USGS) Environmental Health Mission Area's Toxic Substances Hydrology Program and Contaminants Biology Program, and to P.B. by the Institut Universitaire de France. The Froggy platform of the CIMENT infrastructure (ANR grant ANR-10-EQPX- 29-01) provided computing resources, and Pierre Girard contributed his expertise in parallel scientific processing. We thank Sarah Janssen (USGS) and Grégoire Pascaud (LIENs) for assistance with Hg and Se measurements, Stéphane Guédron (ISTerre) for valuable discussions on Hg biogeochemistry, and Andrea Foster (USGS) who provided useful comments on the manuscript. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

■ REFERENCES

(1) Villar, E.; Cabrol, L.; Heimbürger-Boavida, L. E. Widespread microbial mercury methylation genes in the global ocean. *Environ. Microbiol. Rep.* **2020**, *12*, 277−287.
(2) Harris, H. H.; Pickering, I. J.; George, G. N. The chemical form of mercury in fish. *Science* **2003**, *301*, 1203.
(3) Kuwabara, J. S.; Arai, Y.; Topping, B. R.; Pickering, I. J.; George, G. N. Mercury speciation in piscivorous fish from mining-impacted reservoirs. *Environ. Sci. Technol.* **2007**, *41*, 2745−2749.
(4) George, G. N.; MacDonald, T. C.; Korbas, M.; Singh, S. P.; Myers, G. J.; Watson, G. E.; O'Donoghue, J. L.; Pickering, I. J. The chemical forms of mercury and selenium in whale skeletal muscle. *Metallomics* **2011**, *3*, 1232−1237.
(5) Korbas, M.; O'Donoghue, J. L.; Watson, G. E.; Pickering, I. J.; Singh, S. P.; Myers, G. J.; Clarkson, T. W.; George, G. N. The chemical nature of mercury in human brain following poisoning or environmental exposure. *ACS Chem. Neurosci.* **2010**, *1*, 810−818.
(6) Manceau, A.; Enescu, M.; Simionovici, A.; Lanson, M.; Gonzalez-Rey, M.; Rovezzi, M.; Tucoulou, R.; Glatzel, P.; Nagy, K. L.; Bourdineaud, J.-P. Chemical forms of mercury in human hair reveal sources of exposure. *Environ. Sci. Technol.* **2016**, *50*, 10721−10729.
(7) Arai, T.; Ikemoto, T.; Hokura, A.; Terada, Y.; Kunito, T.; Tanabe, S.; Nakai, I. Chemical forms of mercury and cadmium accumulated in marine mammals and seabirds as determined by XAFS analysis. *Environ. Sci. Technol.* **2004**, *38*, 6468−6474.
(8) Nakazawa, E.; Ikemoto, T.; Hokura, A.; Terada, Y.; Kunito, T.; Tanabe, S.; Nakai, I. The presence of mercury selenide in various tissues of the striped dolphin: evidence from μ-XRF-XRD and XAFS analyses. *Metallomics* **2011**, *3*, 719−725.
(9) Lailson-Brito, J.; Cruz, R.; Dorneles, P. R.; Andrade, L.; Azevedo, A. d. F.; Fragoso, A. B.; Vidal, L. G.; Costa, M. B.; Bisi, T. L.; Almeida, R.; Carvalho, D. P.; Bastos, W. R.; Malm, O. Mercury-selenium relationships in liver of Guiana dolphin: The possible role of Kupffer cells in the detoxification process by tiemannite formation. *PLoS One* **2012**, *7*, No. e42162.
(10) Sakamoto, M.; Itai, T.; Yasutake, A.; Iwasaki, T.; Yasunaga, G.; Fujise, Y.; Nakamura, M.; Murata, K.; Man Chan, H.; Domingo, J. L.; Marumoto, M. Mercury speciation and selenium in toothed-whale muscles. *Environ. Res.* **2015**, *143*, 55−61.
(11) Gajdosechova, Z.; Lawan, M. M.; Urgast, D. S.; Raab, A.; Scheckel, K. G.; Lombi, E.; Kopittke, P. M.; Loeschner, K.; Larsen, E. H.; Woods, G.; Brownlow, A.; Read, F. L.; Feldmann, J.; Krupp, E. M. In vivo formation of natural HgSe nanoparticles in the liver and brain of pilot whales. *Sci. Rep.* **2016**, *6*, 34361.
(12) Manceau, A.; Gaillot, A. C.; Glatzel, P.; Bustamante, P. In vivo formation of HgSe nanoparticles and Hg-tetraselenolate complex from methylmercury in seabird − Implications for the Hg-Se antagonism. *Environ. Sci. Technol.* **2021**, in press, DOI: 10.1021/acs.est.0c06269.
(13) Huggins, F. E.; Raverty, S. A.; Nielsen, O. S.; Sharp, N. E.; Robertson, J. D.; Ralston, N. V. C. An XAFS investigation of mercury and selenium in Beluga whale tissues. *Environ. Bioindic.* **2009**, *4*, 291−302.
(14) Guédron, S.; Grangeon, S.; Jouravel, G.; Charlet, L.; Sarret, G. Atmospheric mercury incorporation in soils of an area impacted by a chlor-alkali plant (Grenoble, France): Contribution of canopy uptake. *Sci. Total Environ.* **2013**, *445*−*446*, 356−364.
(15) Rovezzi, M.; Lapras, C.; Manceau, A.; Glatzel, P.; Verbeni, R. High energy-resolution x-ray spectroscopy at ultra-high dilution with spherically bent crystal analyzers of 0.5 m radius. *Rev. Sci. Instrum.* **2017**, *88*, 013108.
(16) Manceau, A.; Wang, J.; Rovezzi, M.; Glatzel, P.; Feng, X. Biogenesis of mercury-sulfur nanoparticles in plant leaves from atmospheric gaseous mercury. *Environ. Sci. Technol.* **2018**, *52*, 3935−3948.
(17) Bourdineaud, J.-P.; Gonzalez-Rey, M.; Rovezzi, M.; Glatzel, P.; Nagy, K. L.; Manceau, A. Divalent Mercury in Dissolved Organic Matter Is Bioavailable to Fish and Accumulates as Dithiolate and Tetrathiolate Complexes. *Environ. Sci. Technol.* **2019**, *53*, 4880−4891.
(18) Gailer, J.; George, G. N.; Pickering, I. J.; Madden, S.; Prince, R. C.; Yu, E. Y.; Denton, M. B.; Younis, H. S.; Aposhian, H. V. Structural basis of the antagonism between inorganic mercury and selenium in mammals. *Chem. Res. Toxicol.* **2000**, *13*, 1135−1142.
(19) Nagy, K. L.; Manceau, A.; Gasper, J. D.; Ryan, J. N.; Aiken, G. R. Metallothionein-like multinuclear clusters of mercury(II) and sulfur in peat. *Environ. Sci. Technol.* **2011**, *45*, 7298−7306.
(20) Dunham-Cheatham, S.; Mishra, B.; Myneni, S.; Fein, J. B. The effect of natural organic matter on the adsorption of mercury to bacterial cells. *Geochim. Cosmochim. Acta* **2015**, *150*, 1−10.
(21) Poulin, B. A.; Gerbig, C. A.; Kim, C. S.; Stegemeier, J. P.; Ryan, J. N.; Aiken, G. R. Effects of sulfide concentration and dissolved organic matter characteristics on the structure of nanocolloidal metacinnabar. *Environ. Sci. Technol.* **2017**, *51*, 13133−13142.
(22) Korbas, M.; Percy, A. J.; Gailer, J.; George, G. N. A possible molecular link between the toxicological effects of arsenic, selenium

and methylmercury: methylmercury(II) seleno bis(S-glutathionyl) arsenic(III). *J. Biol. Inorg Chem.* **2008**, *13*, 461−470.
(23) Pickering, I. J.; Cheng, Q.; Rengifo, E. M.; Nehzati, S.; Dolgova, N. V.; Kroll, T.; Sokaras, D.; George, G. N.; Arnér, E. S. J. Direct observation of methylmercury and auranofin binding to selenocysteine in thioredoxin reductase. *Inorg. Chem.* **2020**, *59*, 2711−2718.
(24) Glatzel, P.; de Groot, F. M. F.; Manoilova, O.; Grandjean, D.; Weckhuysen, B. M. Range-extended EXAFS at the L edge of rare earths using high-energy-resolution fluorescence detection: A study of La in LaOCl. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2005**, *72*, 014117.
(25) Pushkar, Y.; Yano, J.; Glatzel, P.; Messinger, J.; Lewis, A.; Sauer, K.; Bergmann, U.; Yachandra, V. Structure and orientation of the Mn_4Ca cluster in plant photosystem II membranes studied by polarized range-extended X-ray absorption spectroscopy. *J. Biol. Chem.* **2007**, *282*, 7198−7208.
(26) Bianchini, M.; Glatzel, P. A tool to plan photon-in/photon-out experiments: count rates, dips and self-absorption. *J. Synchrotron Radiat.* **2012**, *19*, 911−919.
(27) Busto, M. E. d. C.; Oster, C.; Cuello-Nuñez, S.; Deitrich, C. L.; Raab, A.; Konopka, A.; Lehmann, W. D.; Goenaga-Infante, H.; Fisicaro, P. Accurate quantification of selenoproteins in human plasma/serum by isotope dilution ICP-MS: focus on selenoprotein P. *J. Anal. At. Spectrom.* **2016**, *31*, 1904−1912.
(28) Gajdosechova, Z.; Mester, Z.; Feldmann, J.; Krupp, E. M. The role of selenium in mercury toxicity - Current analytical techniques and future trends in analysis of selenium and mercury interactions in biological matrices. *TrAC, Trends Anal. Chem.* **2018**, *104*, 95−109.
(29) Ackerman, J. T.; Hartman, C. A.; Eagles-Smith, C. A.; Herzog, M. P.; Davis, J.; Ichikawa, G.; Bonnema, A. Estimating mercury exposure of piscivorous birds and sport fish using prey fish monitoring. *Environ. Sci. Technol.* **2015**, *49*, 13596−13604.
(30) Ackerman, J. T.; Eagles-Smith, C. A.; Herzog, M. P.; Hartman, C. A.; Peterson, S. H.; Evers, D. C.; Jackson, A. K.; Elliott, J. E.; Vander Pol, S. S.; Bryan, C. E. Avian mercury exposure and toxicological risk across western North America: A synthesis. *Sci. Total Environ.* **2016**, *568*, 749−769.
(31) Poulin, B. A.; Manceau, A.; Krabbenhoft, D. P.; Stewart, A. R.; Ward-Deitrich, C.; del Castillo Busto, M. E.; Goenaga-Infante, H.; Bustamante, P. *Mercury and Selenium Chemical Characteristics and Speciation Data of Bird, Fish, and Earthworm Tissues*; U.S. Geological Survey data release, 2020.
(32) Burger, J.; Jehl, J. R.; Gochfeld, M. Selenium:mercury molar ratio in eared grebes (Podiceps nigricollis) as a possible biomarker of exposure. *Ecol. Indic.* **2013**, *34*, 60−68.
(33) Burk, R. F.; Hill, K. E. Regulation of selenium metabolism and transport. *Annu. Rev. Nutr.* **2015**, *35*, 109−134.
(34) Thomas, S. A.; Mishra, B.; Myneni, S. C. B. Cellular mercury coordination environment, and not cell surface ligands, influence bacterial methylmercury production. *Environ. Sci. Technol.* **2020**, *54*, 3960−3968.
(35) Manceau, A.; Bustamante, P.; Haouz, A.; Bourdineaud, J. P.; Gonzalez-Rey, M.; Lemouchi, C.; Gautier-Luneau, I.; Geertsen, V.; Barruet, E.; Rovezzi, M.; Glatzel, P.; Pin, S. Mercury(II) binding to metallothionein in Mytilus edulis revealed by high energy-resolution XANES spectroscopy. *Chem.−Eur. J.* **2019**, *25*, 997−1009.
(36) Palmer, J. H.; Parkin, G. Protolytic cleavage of Hg-C bonds induced by 1-methyl-1,3-dihydro-2H-benzimidazole-2-selone: Synthesis and structural characterization of mercury complexes. *J. Am. Chem. Soc.* **2015**, *137*, 4503−4516.
(37) Rieder, S. R.; Brunner, I.; Horvat, M.; Jacobs, A.; Frey, B. Accumulation of mercury and methylmercury by mushrooms and earthworms from forest soils. *Environ. Pollut.* **2011**, *159*, 2861−2869.
(38) Yamashita, Y.; Yamashita, M. Identification of a novel selenium-containing compound, selenoneine, as the predominant chemical form of organic selenium in the blood of bluefin tuna. *J. Biol. Chem.* **2010**, *285*, 18134−18138.
(39) Labunskyy, V. M.; Hatfield, D. L.; Gladyshev, V. N. Selenoproteins: Molecular pathways and physiological roles. *Physiol. Rev.* **2014**, *94*, 739−777.
(40) Achouba, A.; Dumas, P.; Ouellet, N.; Lemire, M.; Ayotte, P. Plasma levels of selenium-containing proteins in Inuit adults from Nunavik. *Environ. Int.* **2016**, *96*, 8−15.
(41) Chen, C.; Yu, H.; Zhao, J.; Li, B.; Qu, L.; Liu, S.; Zhang, P.; Chai, Z. The roles of serum selenium and selenoproteins on mercury toxicity in environmental and occupational exposure. *Environ. Health Perspect.* **2006**, *114*, 297−301.
(42) Ralston, N. V. C.; Raymond, L. J. Mercury's neurotoxicity is characterized by its disruption of selenium biochemistry. *Biochim. Biophys. Acta, Gen. Subj.* **2018**, *1862*, 2405−2416.
(43) Wang, Q.; Zhang, C. X.; Ren, Y. S.; Yue, W. B.; Shi, L. G.; Lei, F. L. Molecular structure, expression analysis and functional characterization of selenoprotein P (SEPP1) in goat (Capra hircus). *J. Anim. Vet. Adv.* **2012**, *11*, 2898−2904.
(44) Hill, K. E.; Wu, S.; Motley, A. K.; Stevenson, T. D.; Winfrey, V. P.; Capecchi, M. R.; Atkins, J. F.; Burk, R. F. Production of selenoprotein P (Sepp1) by hepatocytes is central to selenium homeostasis. *J. Biol. Chem.* **2012**, *287*, 40414−40424.
(45) Reich, H. J.; Hondal, R. J. Why Nature chose selenium. *ACS Chem. Biol.* **2016**, *11*, 821−841.
(46) Jiang, L.; Ni, J. Z.; Liu, Q. Evolution of selenoproteins in the metazoan. *BMC Genomics* **2012**, *13*, 446.
(47) Yang, J.; Yan, R.; Roy, A.; Xu, D.; Poisson, J.; Zhang, Y. The I-TASSER Suite: Protein structure and function prediction. *Nat. Methods* **2015**, *12*, 7−8.
(48) Khan, M. A. K.; Wang, F. Chemical demethylation of methylmercury by selenoamino acids. *Chem. Res. Toxicol.* **2010**, *23*, 1202−1206.
(49) Asaduzzaman, A. M.; Schreckenbach, G. Degradation Mechanism of Methyl Mercury Selenoamino Acid Complexes: A Computational Study. *Inorg. Chem.* **2011**, *50*, 2366−2372.
(50) Melnick, J. G.; Yurkerwich, K.; Parkin, G. Synthesis, structure, and reactivity of two-coordinate mercury alkyl compounds with sulfur ligands: Relevance to mercury detoxification. *Inorg. Chem.* **2009**, *48*, 6763−6772.
(51) Melnick, J. G.; Parkin, G. Cleaving mercury-alkyl bonds: A functional model for mercury detoxification by MerB. *Science* **2007**, *317*, 225−227.
(52) Yamashita, M.; Yamashita, Y.; Suzuki, T.; Kani, Y.; Mizusawa, N.; Imamura, S.; Takemoto, K.; Hara, T.; Hossain, M. A.; Yabu, T.; Touhata, K. Selenoneine, a novel selenium-containing compound, mediates detoxification mechanisms against methylmercury. Accumulation and toxicity in zebrafish embryo. *Mar. Biotechnol.* **2013**, *15*, 559−570.
(53) Pyykkö, P.; Straka, M. Ab initio studies of the dimers $(HgH_2)_2$ and $(HgMe_2)_2$. Metallophilic attraction and the van der Waals radii of mercury. *Phys. Chem. Chem. Phys.* **2000**, *2*, 2489−2493.
(54) Pyykkö, P. Relativistic effects in chemistry: More common than you thought. *Annu. Rev. Phys. Chem.* **2012**, *63*, 45−64.
(55) Pozzi, C.; Ciambellotti, S.; Bernacchioni, C.; Di Pisa, F.; Mangani, S.; Turano, P. Chemistry at the protein-mineral interface in L-ferritin assists the assembly of a functional (μ^3-oxo)Tris (μ^2-peroxo) triiron(III) cluster. *Proc. Natl. Acad. Sci. U.S.A.* **2017**, *114*, 2580−2585.

https://dx.doi.org/10.1021/acs.est.0c04948
Environ. Sci. Technol. 2021, 55, 1527−1534

Supplementary Information

Demethylation of methylmercury in bird, fish, and earthworm

Alain Manceau[1*], Jean-Paul Bourdineaud[2], Ricardo B. Oliveira[3], Sandra L.F. Sarrazin[3], David P. Krabbenhoft[4], Collin A. Eagles-Smith[5], Joshua T. Ackerman[6], A. Robin Stewart[7], Christian Ward-Deitrich[8], M. Estela del Castillo Busto[8], Heidi Goenaga-Infante[8], Aude Wack[1], Marius Retegan[9], Blanka Detlefs[9], Pieter Glatzel[9], Paco Bustamante[10], Kathryn L. Nagy[11] & Brett A. Poulin[12,13*]

1. Université Grenoble Alpes, ISTerre, CNRS, Grenoble, France
2. Université de Bordeaux, Institut Européen de Chimie et Biologie, CNRS, Pessac, France
3. Universidade Federal do Oeste Pará, LabBBEx, Santarém, Pará, Brazil
4. U.S. Geological Survey, Upper Midwest Water Science Center, Middleton, Wisconsin, USA
5. U.S. Geological Survey, Forest and Rangeland Ecosystem Science Center, Corvallis, Oregon, USA
6. U.S. Geological Survey, Western Ecological Research Center, Dixon Field Station, Dixon, California, USA
7. U.S. Geological Survey, Water Resources Mission Area, Menlo Park, California, USA
8. National Measurement Institute, LGC Limited, Teddington, Middlesex, United Kingdom
9. European Synchrotron Radiation Facility (ESRF), Grenoble, France
10. Université La Rochelle, CNRS, Littoral Environnement et Sociétés, La Rochelle, France
11. University of Illinois at Chicago, Department of Earth and Environmental Sciences, Chicago, Illinois, USA
12. U.S. Geological Survey, Water Resources Mission Area, Boulder, Colorado, USA
13. Department of Environmental Toxicology, University of California Davis, Davis, CA 95616, USA

Supporting information includes 17 Figures, 7 Tables and 43 pages.
Data generated during this study are available at https://doi.org/10.5066/P96NP376.

Table of Contents

S1. Methods

S.1.1. Chemical analysis

The total Hg concentration of the blood sample of the Clark's grebe was quantified on a wet weight basis (1.70 ± 0.17 mg Hg kg^{-1} wet weight), and converted to a dry weight basis (7.04 ± 0.70 mg Hg kg^{-1} dry weight) using the average moisture content of grebe blood (75.9%).[1] Total Hg and MeHg were quantified at the U.S. Geological Survey Mercury Research Laboratory (Middleton, Wisconsin) on tissues on a dry weight basis. Tissues were processed using a nitric acid (4.5 M) extraction and heated at 55 °C for 8 hours. First, MeHg was quantified on an aliquot of extracts by aqueous phase ethylation, trapping on Tenax® (Buchem B.V.), isothermal gas chromatography separation, and cold vapor atomic fluorescence spectroscopy (CVAFS) detection using a Brooks Rand MERX-M following U.S. Environmental Protection Agency Method 1630. Second, tissue extracts were further oxidized with sodium persulfate (2% of final volume) overnight and brominated to a final volume of 10% bromine monochloride (BrCl). Oxidized extracts were measured for total Hg via stannous chloride reduction and gold amalgamation coupled to CVAFS detection using a Brooks Rand MERX-T following U.S. Environmental Protection Agency Method 1631. The standard deviation between digestion triplicates for MeHg was ≤ 5.2% (n=3), the recovery of MeHg from IAEA-436 reference (n=9; certified at 0.200 ± 0.010 mg MeHg kg^{-1} dry weight) averaged 92%, and the recovery of MeHg from quality control spikes averaged 99% (n=17). The standard deviation between digestion triplicates for total Hg was ≤ 8.6% (n=4), the recovery of Hg from IAEA-436 reference (n=9; certified at 0.220 ± 0.010 mg Hg kg^{-1} dry weight) averaged 98%, and the recovery of Hg from quality control spikes averaged 99% (n=14). The total error in percent MeHg (%MeHg) by chemical analysis was estimated by propagation of uncertainty associated with MeHg and total Hg concentrations (assigned at 10% for each measurement).

Total Hg was quantified on the < 2 mm soil fraction (Site S1 and S2) using a DMA-80 Hg analyzer (Milestone Dual-cell). The detection limit was 0.003 ng Hg and accuracy was confirmed by analysis of the IAEA-436 reference from the International Atomic Energy Agency. The established value is 4.19 ± 0.36 mg kg^{-1} dry weight and the determined value was 3.96 ± 0.13 mg Hg kg^{-1} dry weight (n = 30).

Selenium concentrations were determined on all samples by isotope dilution-hydride generation-inductively coupled plasma-mass spectrometry (ID-HG-ICP-MS)[2, 3] at the U.S. Geological Survey (Menlo Park, California), and on three tissues from the Clark's grebe also at University of La Rochelle (muscle, kidneys, liver). At the U.S. Geological Survey, quality assurance and quality control were

assessed by procedural blanks (run in triplicate), duplicate digestions (n = 10), and certified reference materials that spanned a range of sample matrices (run in triplicate with each run). Se recoveries for certified reference materials averaged 106% (n = 6) for National Institute of Science and Technology (NIST2976) mussel tissue, 102% (n = 6) for National Research Council Canada (NRCC) dogfish muscle (DORM2), 96% (n = 6) for NRCC dogfish liver (DOLT3), 93% (n=3) for NRCC lobster hepatopancreas (TORT3), and 115% (n = 3) for NRCC marine sediment (MESS-3). At the University of La Rochelle, tissues were digested in duplicate (Clark's grebe muscle and kidneys) and triplicate (Clark's grebe liver) in ultrapure HNO_3 in polytetrafluoroethylene reactors with a MARS-6 microwave. Sample digests were analyzed with an Agilent 7500cx ICP-MS with Octopole Reaction System. Certified reference materials included European Reference Material ERM-CE278k mussel tissue (recovery of 108%, n=1) and NRCC dogfish liver (DOLT5) (recovery of 101%; n=1). Analytical calibration blanks quantifying instrument noise and procedural blanks were subtracted from sample concentrations. Procedural blanks were always less than 10% of sample concentrations and averaged 0.010 mg kg^{-1} dry weight (based on an average sample mass of 10 mg) and precision for samples run in duplicate averaged 3.0% (calculated as the absolute deviation divided by the mean, as a percent). Average Se concentrations are reported on Clark's grebe tissues when data from both laboratories is available. The uncertainty in Se concentrations was assigned at 10% for all measurements, the maximum deviation in recovery of certified reference materials.

Molar ratios of Se to Hg were calculated using total Se concentration and (1) the total Hg concentration ($Se:Hg_{tot}$) and (2) the Hg concentration as $Hg(Sec)_4$ ($Se:Hg(Sec)_4$). The latter was calculated by multiplying the percentage of Hg as $Hg(Sec)_4$ as determined by HR-XANES by the total Hg concentration, and therefore is only reported for tissues where $Hg(Sec)_4$ was quantified. The total error in $Se:Hg_{tot}$ and $Se:Hg(Sec)_4$ values was estimated by propagation of uncertainty associated with Se and total Hg concentrations and HR-XANES spectral fits.Figure S8 presents the $Se:Hg_{tot}$ and $Se:Hg(Sec)_4$ of the bird, fish, and earthworm tissues as a function of the total Hg concentration. Horizontal dashed lines present stoichiometric ratios of 4 (red) and 1 (gray) for reference of ratios of $Hg(Sec)_4$ and HeSe species, respectively. Stoichiometric ratios of $Se:Hg_{tot}$ less than 1 are observed in the feather and muscle of the Clark's grebe and several muscle tissues of fish (*Plagioscion squamosissimus* and *Cichla sp.*, n=4) (Figure S8a). These samples either had no detectable $Hg(Sec)_4$ (Clark's grebe feather, fish muscle tissues) or $Hg(Sec)_4$ was a minor species (Clark's grebe muscle, 11 ± 5% $Hg(Sec)_4$). All other tissues (n=22) exhibited $Se:Hg_{tot}$ greater than 1. Of the tissues where $Hg(Sec)_4$ was quantified, the $Se:Hg(Sec)_4$ was greater than 4 in all samples aside from the Clark's

grebe kidneys (Se:Hg(Sec)$_4$ = 2.1 ± 0.3) and liver tissues (Se:Hg(Sec)$_4$ = 1.3 ± 0.2) (Figure S8b). We attribute the Se:Hg(Sec)$_4$ less than 4 in these two tissues to the presence of $Hg_x(Se,Sec)_y$ clusters/nanoparticulate Hg selenide, as detected in the HR-EXAFS spectra of the Clark's grebe liver (Figure S9). The inverse relationship between Se:Hg(Sec)$_4$ and total Hg concentration (Figure S8b) may reflect a continuum in Hg speciation from Hg(Sec)$_4$ at low total Hg concentration to $Hg_x(Se,Sec)_y$ clusters at high total Hg concentration across the various tissues, but further work is ultimately needed to resolve the biochemical reasons for these observations.

S.1.2. PCR analysis

DNA extraction from soil and earthworm microflora. For the soil sample (site S2), the DNeasy PowerSoil kit (Qiagen, Product 12888) was used to extract bacterial genomic DNA from 0.25 g of soil following the manufacturer's recommendations. For earthworms, individual earthworms were isolated in a plastic box devoid of soil for 24 h and then dissected and the digestive tract opened. With a sterile spatula, the interior of the digestive tract was transferred to a 12 mL tube containing sterile LB medium. Cultures were incubated static at 16 °C for 3 d. After centrifugation, genomic DNA was extracted from the bacterial pellet with a solution of EDTA, Tris, and lysozyme, followed by proteinase K and N-laurylsarkosinate.[4] The genomic DNA was purified by precipitating in 60% ethanol and rinsing several times in 70% ethanol. The DNA was rinsed a final time with 70% ethanol overnight at -20 °C, air dried, and dissolved in a solution of 10 mM Tris (pH 8) and 1 mM EDTA. The concentration of DNA was measured at 260 nm.

Amplification of *hgcA* gene fragments from environmental samples. The *hgcA* gene fragments were amplified from environmental DNA samples using a combination of a forward primer (hgcA-F) targeting the highly conserved putative cap helix motif and a reverse primer (hgcA-912R) targeting a motif between the third and fourth predicted transmembrane helices near the C-terminus[5] (Table S6). These two primers allow the amplification of a product from the *hgcaA* gene around 650 base pairs (bp).[6] The control amplification of an *hgcA* gene fragment (650 bp) from a plasmid containing *hgcA* was confirmed using hgcA-F and hgcA-912R primers and DreamTaq polymerase reagent (Figure S11a); the *hgcA* gene fragment was provided by Hee-Sung Bae (University of Florida). Upon receipt, the plasmid was extracted and purified after alkaline lysis on columns containing a silica membrane (kit NucleoSpin plasmid, Macherey-Nagel, Product 740588.250).

The PCR amplification of environmental samples was performed using identical cycling conditions to that previously described[7] in 20 µL volumes with 10 µL DreamTaq polymerase (Thermo Scientific;

Products K1071), 0.2 mM of dNTP, 0.75 µM of each primer, and the indicated amount of plasmid or genomic DNA. The final concentration of $MgCl_2$ was 1.5 mM. Amplification products were verified by gel electrophoresis on a 1% agarose gel. PCR products between approximately 620-660 bp were excised from the gel and purified on a column containing a silica membrane (kit NucleoSpin Gel and PCR Clean-up, Macherey-Nagel, Product 740609.250). The isolated PCR products were ligated into pGEM T-Easy vector (Promega) overnight at 4 °C according to manufacturer's recommendations. Thereafter, *E. coli* TOP10 cells (Invitrogen Thermo Fisher Scientific, Product C404010), rendered chemically competent for transformation through $CaCl_2$ treatment, were transformed with 2 µL of the ligation mixture. Transformants were selected on ampicillin-containing jellified LB medium (50 µg/mL). Positive recombinants were selected using PCR: 10 colonies were picked out from the Petri dish and each was suspended in a mixture made of 10 µL of sterile milliQ water, 0.5 µL (50 ng) of M13 forward primer, 0.5 µL (50 ng) of M13 reverse primer, and 12.5 µL of 2 x concentrated DreamTaq blend. The thermocycling program was as follows: initial denaturation step at 94 °C for 10 min, 36 cycles at 94 °C for 60 s, 55 °C for 60 s, and 72 °C for 140 s, followed by a final extension step at 72 °C for 10 min. Positive transformants were cultured on LB medium containing ampicillin (50 µg/mL) and the harbored plasmids were purified using the kit NucleoSpin plasmid (Macherey-Nagel, Product 740588.250). The fragment contained in two recombinant plasmids was sequenced using M13 forward and reverse primers (GATC Biotech, Eurofins Genomics Company).

PCR results. Soil from site S2 contained bacterial DNA with the *hgcA* gene with an amplified gene fragment length of approximately 630 bp (Figure S11b, lane 3). After purification of total genomic DNA from earthworm microflora, amplified *hgcA* gene fragments of approximately 630 to 650 bp were identified for earthworms from sites S1 and S2 (Figure S11c). At site S1, the *hgcA* fragments were from anecic earthworm microflora (Figure S11c, panel i, lane 3), and at site S2 they were from both anecic and endogeic earthworm microflora (Figure S11c, panel ii, lanes 3 and 4, respectively).

Only a single *hgcA* sequence was retrieved from the excised and purified bands from both the soil and earthworm microflora samples. The retrieved *hgcA* sequence (633 nucleotides, 211 amino acid residues; Figure S12a) matches *hgcA* sequences belonging to the *Geobacter* genus. Since earthworms ingest soil, the *hgcA* gene-harboring bacteria detected in the gut microflora of earthworms are interpreted to have originated in the soil. At the nucleotide level, the retrieved *hgcA* sequence best matched the *hgcA* sequence of *Geobacteraceae* clone B2_FE06 (Figure S12b), which was amplified from the total bacterial genomic DNA of a temperate forested wetland in southern Sweden.[6] The retrieved *hgcA* sequence contained 437 of 633 identical nucleotides to that of *Geobacteraceae* clone

B2_FE06 (identity percentage of 69%). At the protein level, the amino acid sequence of the retrieved *hgcA* gene best matched a sequence of *Geobacteraceae* isolated from paddy soil in the Guizhou province of southwest China[8]; 43 of 211 amino acid residues are identical (identity percentage of 68%; Figure S12c).

S.1.3. HR-XANES and HR-EXAFS measurement

Freeze-drying a frozen tissue does not change the speciation of chemically-bound metals,[9] which we verified previously on MeHgCys.[10] All freeze-dried samples were pressed into 5.0 mm diameter and 2.5 mm thick pellets, mounted in a polyether ether ketone (PEEK) sample holder sealed with Kapton tape, and maintained in a desiccator until their transfer into the liquid helium cryostat of the beamline. Mercury L_3-edge HR-XANES and HR-EXAFS spectra were measured with high-reflectivity analyzer crystals[11] on beamline ID26 at the European Synchrotron Radiation Facility (ESRF). The storage ring was operated in the 7/8 + 1 filling mode, with 200 mA current. Rejection of higher harmonics and reduction of heat load were achieved with a white beam Pd-coated, flat mirror working under total reflection at 2.5 mrad deflecting angle. The energy of the incoming beam was selected by the 111 reflection of a Si double crystal monochromator, and the beam was focused horizontally by a second Pd-coated mirror and vertically by a third Pd-coated mirror. The flux on the sample was approximately 10^{13} photon/s in a beam footprint of ~700 (H) x 80 (V) µm² full width at half maximum (FWHM). The Hg $L_{\alpha 1}$ ($3d_{5/2} \rightarrow 2p_{3/2}$) fluorescence line was selected using the 555 reflection of five spherically bent (radius = 0.5 m) [11] Si analyzer crystals (diameter = 100 mm) aligned at 81.8° Bragg angle in a vertical Rowland geometry. The diffracted intensity was measured with a Si drift diode detector (SDD) in single photon counting mode. The effective energy resolution, obtained by convoluting the total instrument energy bandwidth (spreads of the incident and emitted rays) and the $3d_{5/2}$ core-hole width from the $L_{\alpha 1}$ line was about 3.0 eV, compared to an intrinsic line broadening of about 6.1 eV in conventional fluorescence yield measurement with a solid-state detector (conventional XANES and EXAFS).

The incident energy was scanned from 12260 eV to 12360 eV in 0.2 eV steps in HR-XANES mode, and the spectra were normalized to unity at E = 12360 eV. The incident energy was scanned from 12210 eV to 13050 eV in 2.0 eV steps in HR-EXAFS mode. The stability in energy of the incident beam was monitored by measuring frequently a fresh MeHgCys reference. The photon energy is referenced to the maximum of the near-edge peak of MeHgCys at 12279.8 eV. The precision of the calibrated spectra is 0.1 eV. Spectra were collected at a temperature of 10-15 K and a scan time of 15

s to reduce exposure, and repeated at different pristine positions on the sample to increase the signal-to-noise ratio. Radiation damage, which occurs when the beam is positioned on the same spot for too long, depends on the chemical form of Hg and, for a given form, varies with the kinetic energy of the photoelectron (i.e., energy of the incident beam). Bound states, probed at low kinetic energy, are more sensitive to radiation exposure than unbound states in the continuum. This dependence is illustrated in Figure S17 with MeHgCys, which is the most radiation sensitive Hg species. $Hg(SR)_2$ is more stable, and nanoparticulate β-HgS and HgSe resist damage during long exposure time.

S.1.4. HR-XANES and HR-EXAFS data analysis and accuracy of the fit results

The number of Hg species was determined by principal component analysis,[12, 13] and their nature subsequently by iterative transformation factor analysis (ITFA[14]) for the $Hg(Sec)_4$ species, and by target transformation analysis[15] for the MeHgCys and $Hg(SR)_2$ species. Target transformation analysis was conducted with reference spectra from a large dataset of Hg minerals (α-HgS, β-HgS, $\beta\text{-HgS}_{NP}$, HgSe, $HgSe_{NP}$), and Hg(II) and MeHg model complexes described previously (Ref.[10] and references therein). All reference spectra were considered as a basis for identification, but only diagnostic spectra are discussed in the article. The proportions of the Hg species were evaluated using least-squares fitting (LSF) of the data with linear combinations of the Clark's grebe breast feather spectrum, the $Hg(Sec)_4$ spectrum, and the spectrum of the Hg(L-glutathione)$_2$ complex ($Hg(GSH)_2$) at pH 7.4.[16, 17] In $Hg(GSH)_2$, Hg(II) is bonded approximately linearly to the cysteinyl sulfur atoms at 2.33 Å from the two γ-Glu-Cys-Gly peptides and surrounded in trans-equatorial position by a carboxyl oxygen from a Gly residue at 2.62 Å, the backbone carbonyl oxygen at 2.88 Å from the Gly-Cys peptide bond (Gly−NH−CO−Cys) of the same GSH molecule, and an amide group (−NH) at 3.01 Å from the second GSH molecule.[18] This reference is a good proxy for the $Hg(SR)_2$ coordination in biota.[18] The precision of estimation of a fit component was estimated to be equal to the variation of its value when the fit residual (*NSS*) was increased by 20%. *NSS* is the normalized sum-squared difference between two spectra expressed as $\Sigma[(y_2\text{-}y_1)^2]/\Sigma(y_2^2)$. The principal component analysis (PCA) and linear fit programs from beamline 10.3.2 at the Advanced Light Source were used.[12, 19]

The accuracy of spectral fit results was evaluated by comparing the percentage and concentration of MeHg obtained by chemical analysis and HR-XANES. The relative abundance of MeHg as determined by HR-XANES ($\%MeHgCys_{XANES}$) and chemical analysis ($\%MeHg_{Chem}$) show good agreement between the two independent methods (Figure S4a), demonstrating a high level of accuracy in spectral fit results. In Figure S4a, a generic Y-axis error bar represent the precision of HR-XANES

spectral fits and the X-axis error bar represent the total propagation of error in $\%MeHg_{Chem}$ measurements. Across the tissues a moderately lower percent MeHg was observed by chemical analysis compared to HR-XANES. We attribute this discrepancy primarily to the lower recovery of MeHg in tissue digests during the ethylation step of the analysis (average of 92% MeHg recovery for IAEA-436, *n*=9). In contrast, HR-XANES analysis does not require sample modification prior to measurement, is sensitive to all Hg atoms in a tissue sample, MeHgCys has a distinctive spectral signature with an intense near-edge peak, and the percentage MeHgCys is obtained directly from a spectral fit. To assess the uncertainty of estimation further, the concentrations of MeHg determined by HR-XANES ($[MeHgCys]_{XANES}$ = %MeHgCys x $[Hg]_{Tot}$) and chemical analysis ($[MeHg]_{Chem}$) were compared (Figure S4b and Table S7). Good agreement was observed between $[MeHgCys]_{XANES}$ and $[MeHg]_{Chem}$, demonstrating also high certainty in the quantification of MeHgCys by the two methods.

The HR-EXAFS spectrum from the Clark's grebe liver was analyzed with WinXAS[20] using amplitude and phase shift functions generated by FEFF 7.0.2[21] and HgSe as structure model[22] (Table S4). The amplitude reduction factor S_0^2 was fixed to 0.9. EXAFS lacks sensitivity to the C atom at about 2.07 Å from the methyl group of MeHgCys and misses Hg atoms at about 4.46 Å as a result of the high disorder of the $Hg_x(Se,Sec)_y$ clusters. Thus, the coordination number (*CN*) of Hg at 4.46 Å is an apparent number.

S.1.5. Protein extraction and analysis

Protein extraction from Clark's grebe tissues. To assess if Hg and Se were associated with SelP, a protein extraction was performed on the Clark's grebe liver and muscle tissues.[23] In triplicate, 0.1 g of lyophilized, homogenized tissue was mixed with 0.4 g of protein extraction buffer (50 mM TRIS-HCl, CAS 1185-53-1; 0.01 g/g SigmaFast Protease Inhibitor Cocktail, Sigma Aldrich) in 1.5 mL Eppendorf tubes. Procedural extraction blanks were performed in triplicate. At 4 °C, samples were homogenized (Omni Tissue Homogenizer with plastic disposable tips) using 5 repeated cycles of a 10 second homogenization and 50 second pause. Protein extracts were separated from residual tissues by centrifugation at 10,000 g at 4 °C for 30 minutes. The supernatant was immediate frozen at -80 °C and shipped at -80 °C to LGC Group (Teddington, United Kingdom) for selenoprotein analysis.

Speciation of Hg and Se in Clark's grebe tissues. Protein extracts from Clark's grebe liver tissue (*n* = 3) and muscle tissue (*n* = 3) and extraction blanks (*n* = 3) were screened for the co-elution of Hg and Se with selenoproteins by double affinity (AF) high performance liquid chromatography-inductively coupled plasma mass spectrometry (HPLC-ICPMS[3, 24]). All samples were measured in parallel in the

same analytical run. Each sample was measured in duplicate. The double AF-HPLC chromatographic system achieves separation of three major Se protein species by the combination of two AF columns (HiTrap Heparin High Performance, 1mL; Hi Trap Blue High Performance, 1 mL) that retain SelP and SeAlb, respectively. Peak 1 elutes forms of Se not retained on the columns, including glutathione peroxidase 3 (GPX3), inorganic Se species, and LMW Se species (retention time of approximately 3.3 minutes). Peak 2 elutes at the retention time of SelP (approximately 19.3 minutes), known to have high affinity to heparin. Peak 3 elutes at the retention time of Se-albumin (Se-Alb; approximately 28.0 minutes). The plasma reference material NIST SRM 1950 was used as the quality control material for chromatographic quality assessment (Se species elution distribution and order), because of the lack of any certified material for Se-species in tissue. It should be noted that thioredoxin reductase (TrxR) elutes at the same retention time as SelP, based on the analysis of a protein standard of TrxR from rat liver. Se species were eluted using ammonium acetate buffers (0.05 and 1.5 M, pH 7, 0.5 mL min^{-1}). Eluent from the AF-HPLC was introduced to a collision cell ICP-MS/MS for measurement of Se and Hg (8800 ICP-QQQMS, Agilent Technologies, UK). The ICP-MS/MS was operated in oxygen mode, and therefore Se was determined by measurement of ^{80}Se as ^{96}SeO. Mercury was measured by monitoring ^{202}Hg. Relative elemental percentages were calculated using the peak area distribution of Peaks 1-3. Hg mass balance of tissue extracts associated with Se species was not performed.

Good repeatability in Hg and Se chromatograms of triplicate tissue extracts was observed. An example of Hg and Se chromatograms of the Grebe liver and muscle extracts are presented in Figure 4. In the liver and muscle tissues, the majority of detected Hg co-eluted with Peak 2 (SelP + TrxR; 87.8 ± 0.4% and 84.3 ± 0.7%, respectively), and the remainder of Hg co-eluted with Peak 1 (GPX3, inorganic Se species, LMW Se species; 11.5 ± 0.4% and 13.7 ± 0.6% respectively). Similarly, Se chromatograms show that Se co-eluted with Peak 1 (71.4 ± 0.7% and 62.5 ± 2.4% in liver and muscle tissue extracts, respectively) and Peak 2 (27.3 ± 0.8% and 35.5 ± 2.3%, respectively). No considerable Hg or Se in either tissue extract was observed to co-elute with Peak 3 (Se-Alb; ≤ 2%). The Se and Hg chromatograms of extraction blanks contained only minor peaks for Hg also observed in the eluent blank (signal ≤ 3% of tissue extracts), confirming that Hg and Se chromatograms of tissue extracts were free of contamination. In summary, this speciation analysis shows that Hg co-elutes primarily with SelP and TrxR and minorly with GPX3 in Grebe liver and muscle tissues, where HR-XANES quantified 86% and 11% of Hg as $Hg(Sec)_4$ (Table S5). Based on biostructural considerations, formation of $Hg(Sec)_4$ is incompatible with TrxR because this homodimeric protein containing a single Sec residue per subunit[25], therefore binds Hg linearly.[26] Similarly, the minor elution of Hg with Peak

1 is considered unlikely to be a result of Hg association with GPX3, given that GPX3 is a tetramer containing one Sec residue in each subunit[27] too far apart to form a $Hg(Sec)_4$ complex. In addition, the HR-XANES spectra of MeHgSec and $Hg(Sec)_2$ species are clearly distinguished from MeHgCys and $Hg(GSH)_2$ (Figure S2c, d), further supporting that co-elution of Hg and Se with Peak 2 is likely as $Hg(Sec)_4$.

S.1.6. SelP amino acid sequence

Crested grebe (*Podiceps cristatus*). The partial protein sequence has been established from GenBank: KL268852.1 by joining two translated nucleotides fragments from locations 15116-15235 and 16132-16350. This incomplete sequence is also found under the protein accession number GenBank: KFZ63143.1. The trace of the first nucleotidic fragment is underlined below.

Incomplete SelP protein:

RCGRLVYHLGLPYSFLSFQYVEESIKIAYCENKCGNCSYTEPAIDDICENITKTADEKLAEPEPKPTGQHSHH
HNLHRHRHHHHHHHREGSRHSKNENHQNSSETQRHHPHSGHNRHDHTGSHEQVDTVPPGESVEISQDKKLUKK
GKTSCKNQLTUNWQTASDSASSSUCCHURHLLFEELGNSVTUQCRGALPNSCRUHGQLSAEDITESUQURLLT
AAUQSPAAGASETSDTUQUQEKARNUAUKTN

C-terminal region used to model the Hg-binding site

DTUQUQEKARNUAUKTN

Zebrafish (*Danio rerio*). The SelP sequence has been established from the protein accession number GenBank: NP_840082.3.

SelP protein:

MWKALSLTLALCLLVGCSAESETEGARCKLPPEWKVGDVEPMKNALGQVTVVAYLQASULFCLEQASKLNDLL
LKLEKQGYPNIAYMVVNNREERSQRLHHLLQERLLNITLYAQDLSQPDAWQAVNAEKDDILVYDRCGRLTYHL
SLPYTILIHPHVEEAIKHTYCDRICGECSLESSAQLEECKKATEEVNKPVEEEPRQDHGHHEQGHHEHQGEAE
RHRHGHHHPHHHHHHHRGQQQVDVDQQVLSQVDFGQVAVETPMMKRPUAKHSRUKVQYSUQQGADSPVASUCU
HURQLFGGEGNGRVAGLUHCDEPLPASUPUQGLKEQDNHIKETUQURPAPPAEUELSQPTUVUPAGDATUGUR
KK

C-terminal region used to model the Hg-binding site

PTUVUPAGDATUGURKK

Fish, birds, and mammals. We performed BLAST analyses with known SelP sequences in fish, birds, and mammals as queries to search protein databases for distant SelP homologs. The sequences

identified served as new search entries for SelP homologs in nucleotide sequence databases. Multiple alignments of SelP sequences (Figure S13) were carried out using the DNAMAN software.

S2. Supplementary figures

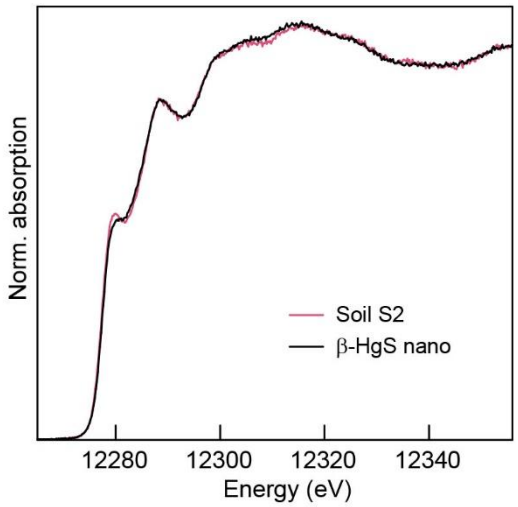


Figure S1. Mercury L_3-edge HR-XANES spectra of the <63 µm fraction of soil S2 (27.6 mg Hg kg^{-1} dw) and the clay-size fraction (0.1 – 2.0 µm) from the A horizon soil (191 mg Hg kg^{-1} dw) studied in Ref.[28]. Hg is speciated as nanoparticulate β-HgS in the A horizon sample.

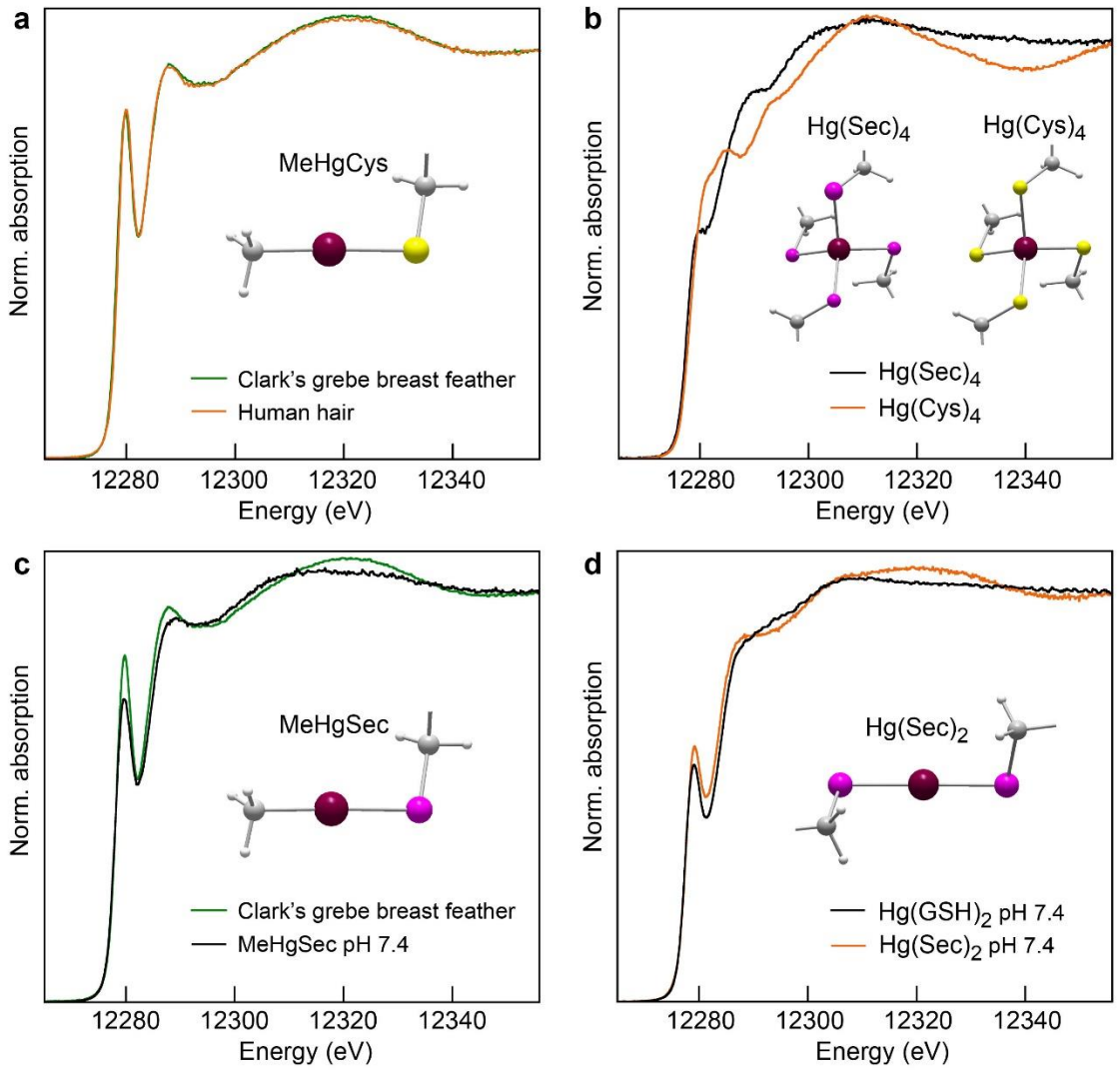


Figure S2. a) HR-XANES spectra of MeHgCys in human hair and Clark's grebe breast feather. The human hair spectrum corresponds to sample 1MeHg13.8 ([Hg] = 13.8 ppm) from Figure 3 of Ref.[29]. b-d) HR-XANES spectra of cysteine and selenocysteine Hg complexes illustrating the high sensitivity of HR-XANES to the coordination and ligation of Hg.

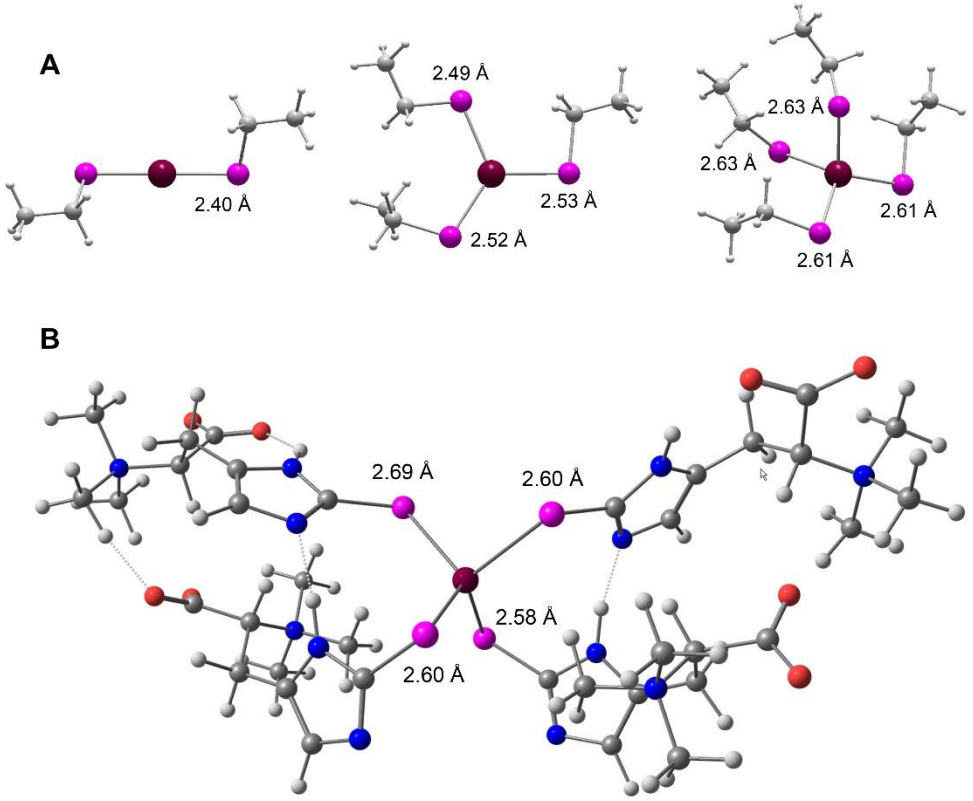


Figure S3. Geometry optimization of the Hg di-, tri-, and tetraselenolate complexes (modeled with ethylselenol) (a) and Hg tetraselenoneine complex ($Hg(selenoneine)_4$) (b) performed at the second-order Møller-Plesset perturbation theory (MP2)[30] level using ORCA 4.2.1[31] and a computational scheme tested previously on the modeling of the structure and stability of monomeric Hg-thiolate complexes[29, 32]. The Ahlrichs polarized def2-TZVP (Hg, Se) and def2-SVP (C, H) basis sets were used in combination with the auxiliary def2-TZVP/C[33, 34] Coulomb and exchange fitting basis to accelerate the MP2 calculations with the resolution of identity (RI) approximation[35, 36]. For Hg atoms, scalar relativistic effects were accounted with an effective core potential (ECP)[37], as obtained from the Stuttgart pseudopotential library (60 core electrons, l_{max}=3). The valence basis set used in connection with the ECP is based on a (8s8p6d1f)/[6s5p3d1f] contraction scheme. The water solvent was represented with the SMD model[38]. The convergence for the self-consistent field (SCF) calculations was set to 10^{-8} a.u. Dark red, purple, gray, blue, red, and light gray spheres represent Hg, Se, C, N, O, and H, respectively.

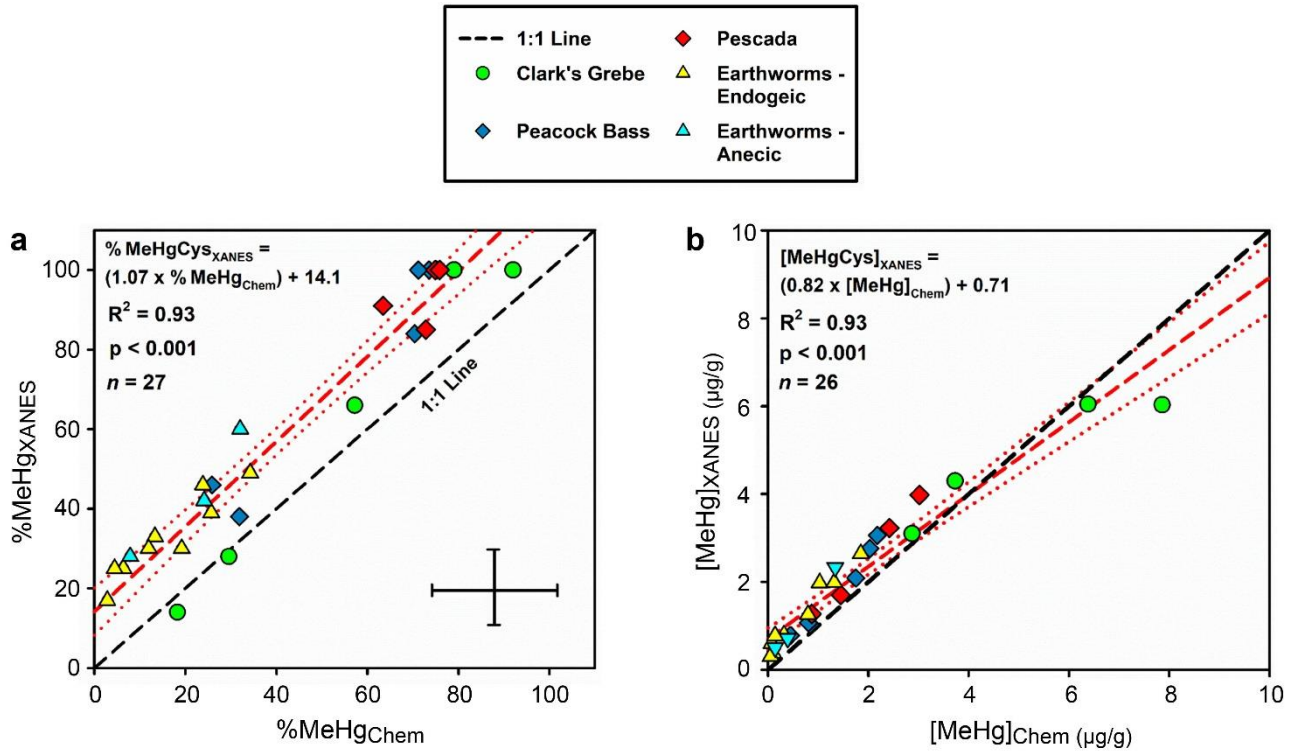


Figure S4. a) The accuracy of spectral fit results was evaluated by comparing the percentage of Hg as MeHg-thiolate complex determined by HR-XANES (%MeHgCys$_{XANES}$) and percentage of MeHg determined by chemical analyses (%MeHg$_{Chem}$). Maximum uncertainties in measurements are expressed as a generic error bar. b) The uncertainty in spectral fit results was evaluated by comparing the concentrations of MeHg determined by HR-XANES ([MeHgCys]$_{XANES}$ = %MeHgCys$_{XANES}$ x [Hg]$_{Tot}$) and chemical analysis ([MeHg]$_{Chem}$). The dashed black line presents the 1:1 line, the dashed red line presents the linear fit of data, and the dotted red lines present the 95% confidence intervals of the fit.

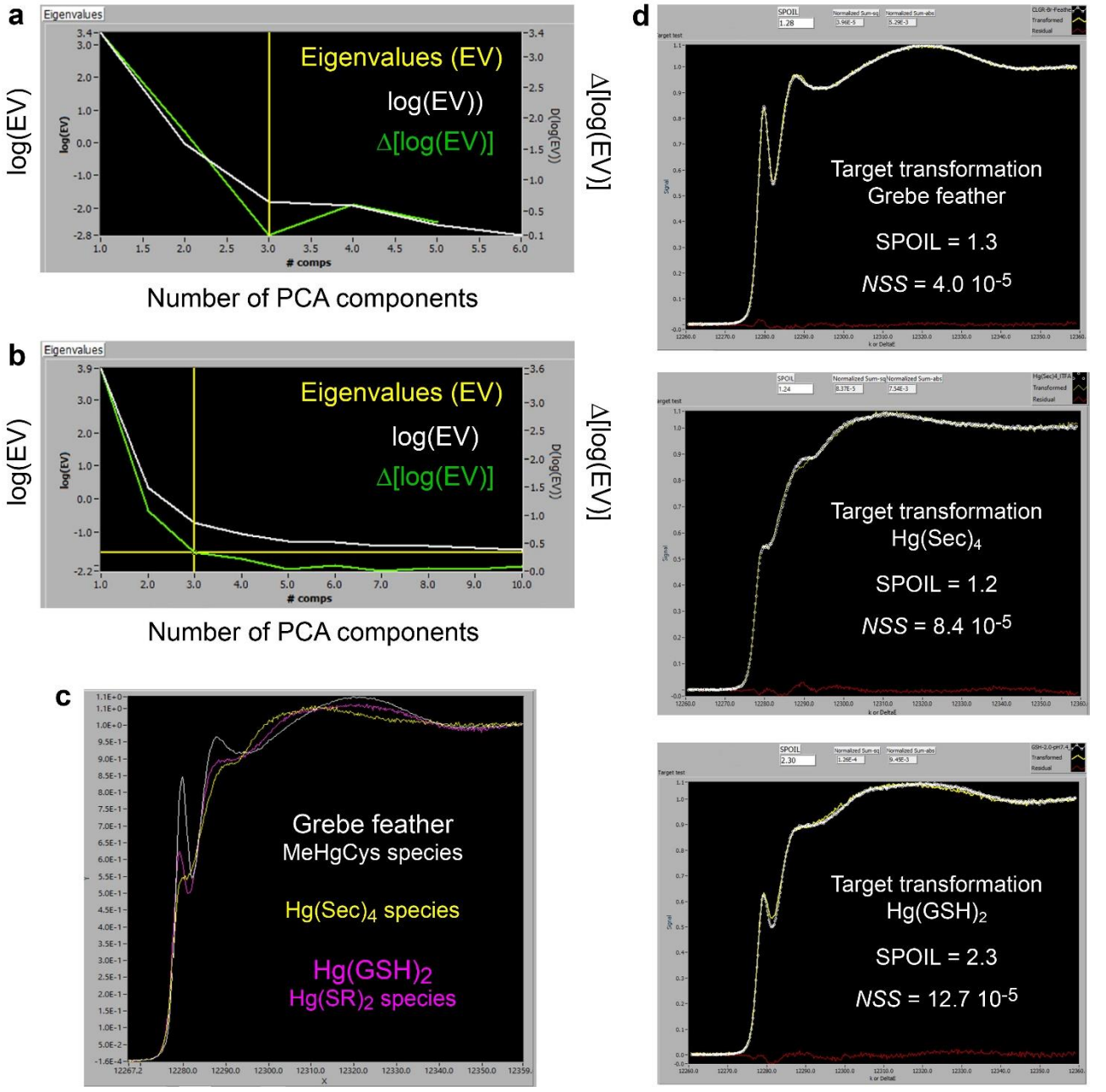


Figure S5. a,b) Estimation by principal component analysis (PCA) of the number of significant abstract components contained in the six HR-XANES spectra from the grebe (a), and the 10 fish + 13 earthworm spectra (b). The curves are a scree plot of the components and the associated eigenvalues (EV) [39]. The variance of each dataset can be accounted for with three components (i.e., Hg species). See ref.[13] for details. **c**, Spectrum of $Hg(Sec)_4$ identified by iterative transformation factor analysis (ITFA[14]), and spectra of MeHgCys and $Hg(SR)_2$ identified by target transformation analysis[15] of a large library of Hg organic and inorganic model compounds[10, 17, 18, 28, 29, 32]. The $Hg(SR)_2$ species was identified previously in fish contaminated by inorganic Hg(II)[18]. d) Target transformed spectra of the three Hg species. The quality of the reconstruction is evaluated with the *SPOIL* criteria[15] and *NSS* value. *SPOIL* measures the degree to which replacing an abstract component with the candidate spectrum would increase the fit error. This is a non-negative dimensionless number for which values

< 1.5 are considered excellent, 1.5-3 good, 3-4.5 fair, 4.5-6 poor, and > 6 unacceptable. See ref.[12] for details.

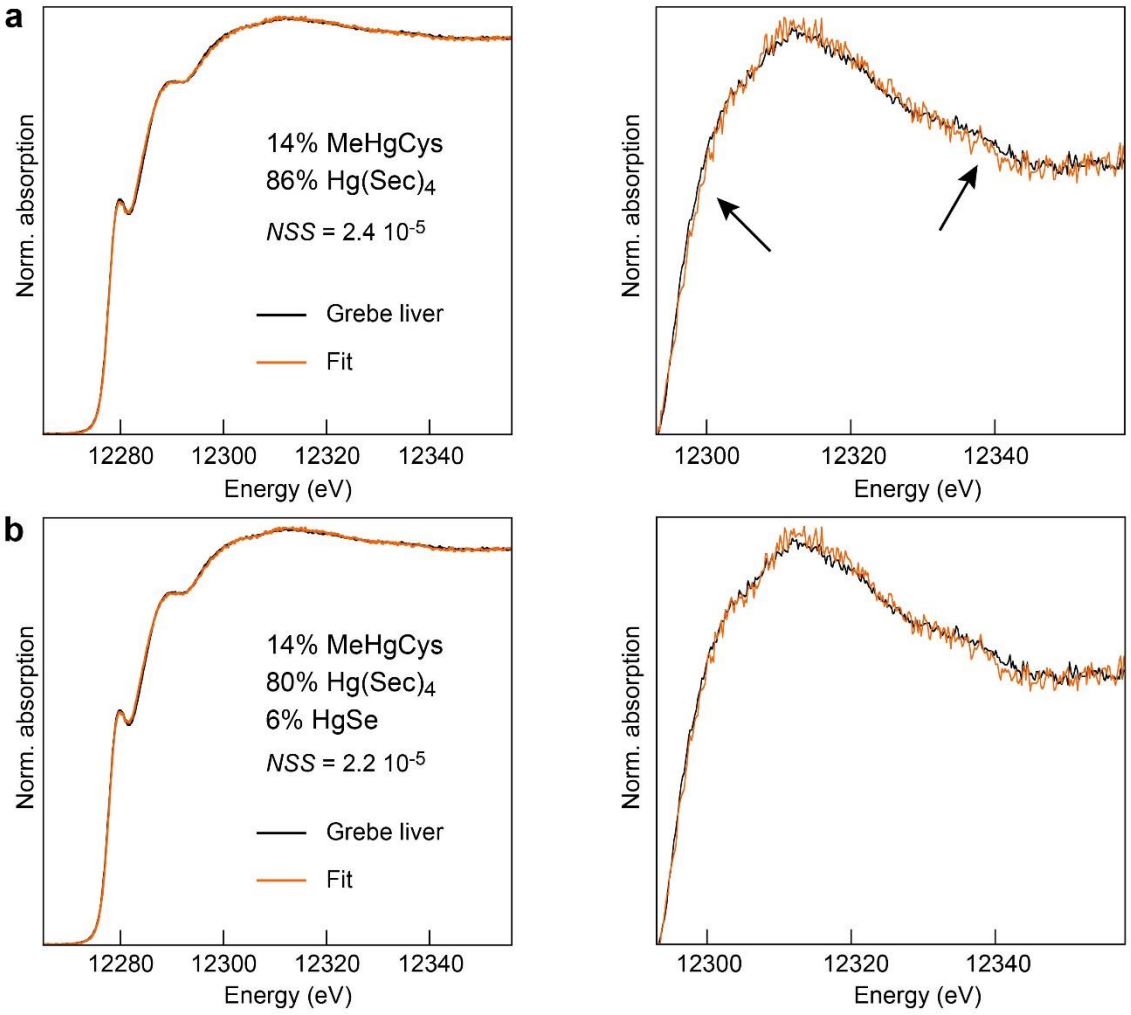


Figure S6. Least-squares fit of the liver HR-XANES spectrum of the Clark's grebe without HgSe (a) and with HgSe (b). The spectra on the right are an enlargement of the top-edge region where the HgSe reference gives a distinctive signal at the energies indicated with arrows. The fit lacks intensity at the diagnostic energies in the absence of the HgSe reference. Including HgSe provides a better fit of the entire grebe liver spectrum including in the diagnostic region where the data and reconstructed spectrum are now statistically superimposed.

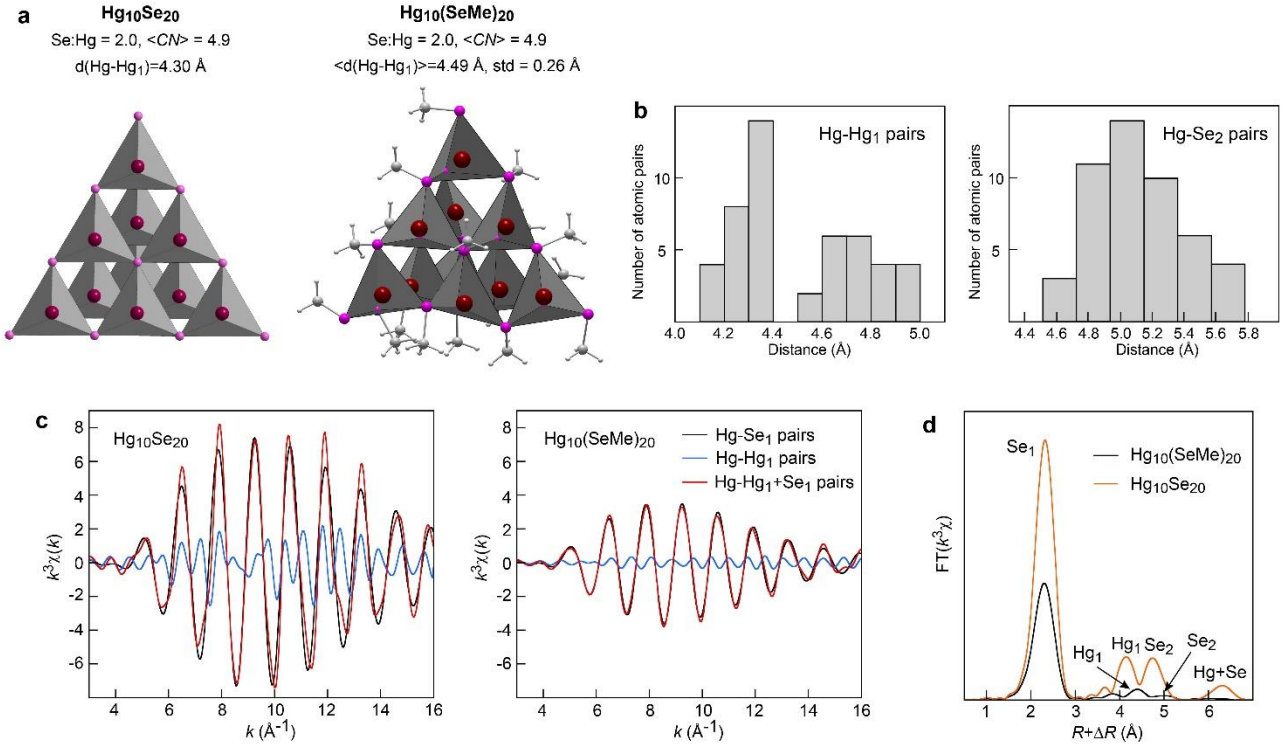


Figure S7. a) Polyhedral representation of a $Hg_{10}Se_{20}$ portion of the HgSe structure[22], and the $Hg_{10}(Sec)_{20}$ cluster optimized geometrically. The $Hg_{10}(Sec)_{20}$ cluster has a predicted first shell Hg-Hg_1 distance of 4.49 Å on average, close to the HR-EXAFS distance of 4.46 Å in the grebe liver (Table S4), and 0.19 Å longer than the crystallographic distance in HgSe (4.30 Å). The optimization was performed at the PBE0-D3 DFT quantum level with def2-TZVP (Hg, Se) and def2-SVP (C, H) basis sets using ORCA 4.2.1[31]. The water solvent was represented with the SMD model[38]. Sec residues were modeled with methaneselenolate (CH_3Se^-). Dark red, Hg; purple, Se; gray, C, light gray, H. b) Population histograms of the predicted first shell Hg-Hg_1 and second shell Hg-Se_2 distances in $Hg_{10}(Sec)_{20}$. Interatomic distances are counted in intervals of 0.1 Å (Hg-Hg_1) and 0.2 Å (Hg-Se_2). The distributions of the interatomic distances are non-Gaussian with considerable asymmetry toward long distances. The Hg-Hg_1 and Hg-Se_2 distances partly overlap, whereas they are separated by 5.04 – 4.30 = 0.74 Å in HgSe. c,d) Calculated EXAFS photoelectron waves using FEFF7[40] of the Hg-Hg_1 plus Hg-Se_1 pairs for $Hg_{10}Se_{20}$ and $Hg_{10}(Sec)_{20}$ and Fourier transforms (radial structure function). The Hg-Hg_1 pair is extremely weak on the radial structure function, as a result of the large dispersion and asymmetry of the Hg-Hg_1 distances and the Hg-Hg_1/Se_2 overlap. Few Hg-Hg_1 pairs are detected, despite the presence, on average, of 4.9 Hg_1 atoms around each Hg atom in the Hg-selenolate cluster. The Hg-Se_1 bond lengths are also unequal and asymmetrical in $Hg_{10}(Sec)_{20}$.

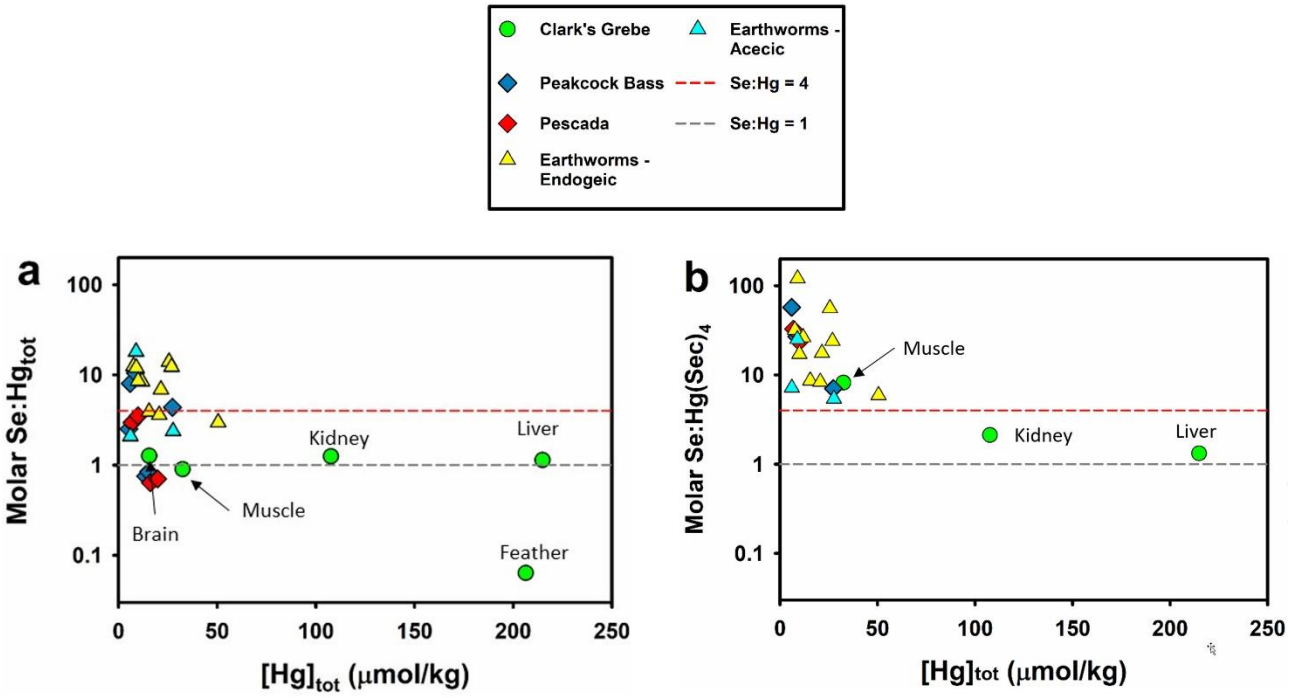


Figure S8. Molar ratios of Se to total Hg ($Se:Hg_{tot}$) versus Hg_{tot} concentration (a) and Se to Hg as a tetraselenolate complex ($Se:Hg(Sec)_4$) versus Hg_{tot} concentration (b) of the Clark's grebe, fish, and earthworm samples. Horizontal dashed lines present stoichiometric ratios of 4 (red) and 1 (gray) for reference of ratios of $Hg(Sec)_4$ and HeSe species, respectively. Identities of the various Clark's grebe tissues are provided. All tissues are presented in plot (a), whereas only tissues with $Hg(Sec)_4$ quantified by HR-XANES are presented in plot (b). Plot (b) shows that tissues containing $Hg(Sec)_4$ have a $Se:Hg(Sec)_4 \geq 4$ except for the Clark's grebe kidneys ($Se:Hg(Sec)_4 = 2.1 \pm 0.3$) and liver tissues ($Se:Hg(Sec)_4 = 1.3 \pm 0.2$), the latter confirmed to contain $Hg_x(Se,Sec)_y$/ nanoparticulate Hg selenide by HR-EXAFS analysis (Figure S9).

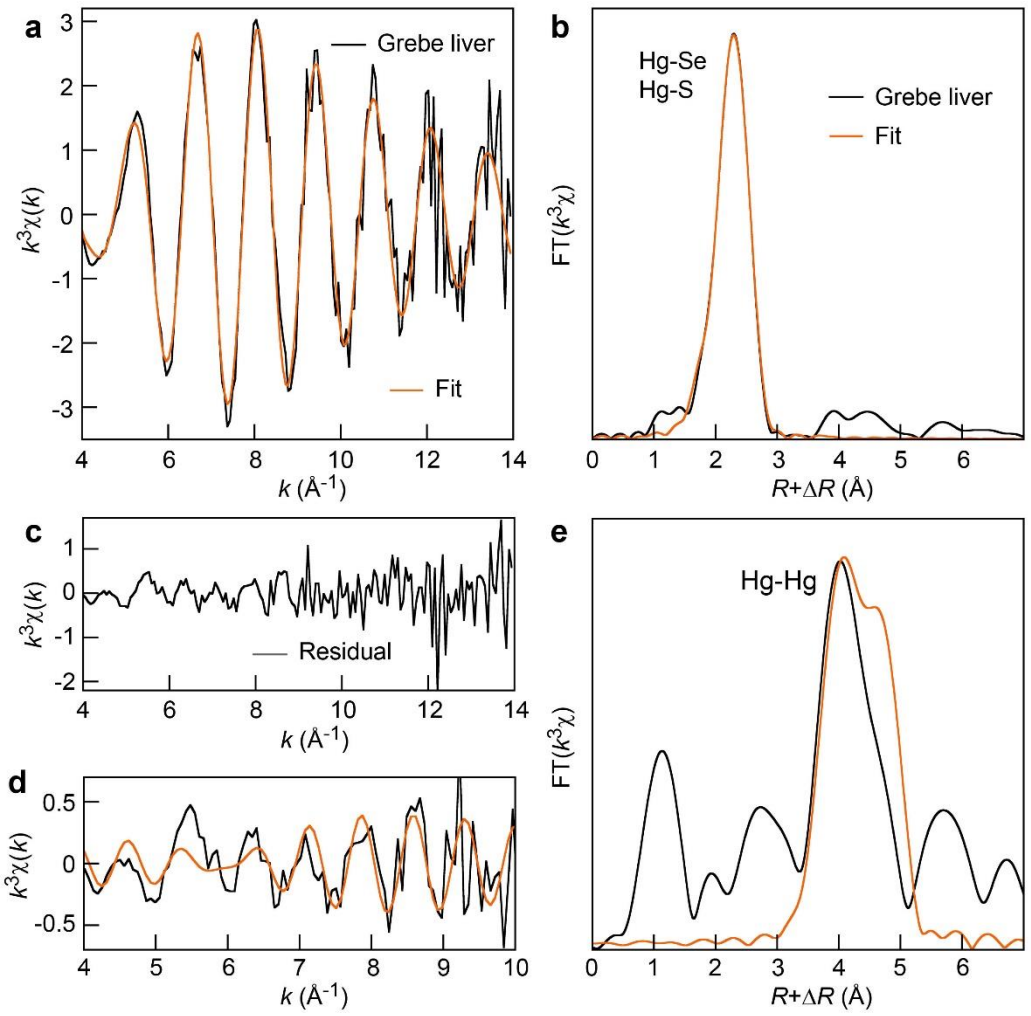


Figure S9. a,b) Fit of the HR-EXAFS spectrum and Fourier transform (radial structure function) for the Clark's grebe liver with Hg-Se and Hg-S pairs in the first coordination shell and no Hg-Hg-pairs in the second coordination shell. **c**, Residual of the HR-EXAFS fit, showing clear evidence for significant spectral structures above noise level that are not accounted for by the Hg-(Se,S) model-fit. c,d) Fit of the HR-EXAFS residual and Fourier transform with Hg-Hg pairs, demonstrating their detection by HR-EXAFS. HR-EXAFS detects 2.1 ± 0.5 Hg atoms at 4.46 ± 0.03 Å with a static disorder of σ = 0.1 Å (Table S4). The distribution of the Hg-Hg distances was considered to be Gaussian in the analysis, which is incorrect when the local structure is highly disordered (e.g., σ ≥ ~0.1 Å[41, 42]) (Figure S7b). An analysis of disordered systems using symmetrical distributions results in "lost" atoms, which means that Hg is in all likelihood surrounded by more than 2.1 Hg atoms. If all the Hg-Hg pairs were from the 6% HgSe detected by HR-XANES (Figure S6b), the number of Hg atoms would be 12 x 0.06 = 0.7. We conclude from this analysis that the majority of the Hg atoms detected by HR-EXAFS are from $Hg_x(Se,Sec)_y$ clusters. Their existence is also supported by the Hg-Hg distance of 4.46 ± 0.03 Å, which is 0.16 Å longer than the Hg-Hg distance in HgSe, and in agreement with the predicted average Hg-Hg distances for disordered $Hg_x(Se,Sec)_y$ cluster models (Figures. S7 and S16).

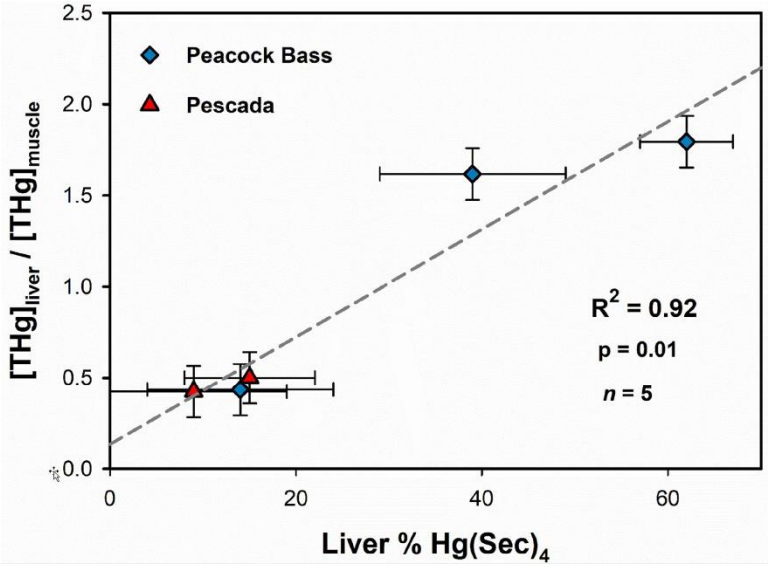


Figure S10. Positive correlation between the total Hg concentration ratio between liver and muscle ($[Hg]_{liver}/[Hg]_{muscle}$) and the percentage of the demethylated fraction of Hg ($\%Hg(Sec)_4$) in the liver tissue of peacock bass. Horizonal error bars present uncertainty in spectral fit data, and vertical error bars present the total propagated error for the ratio of total Hg measurements. The dashed gray line presents the linear fit of data.

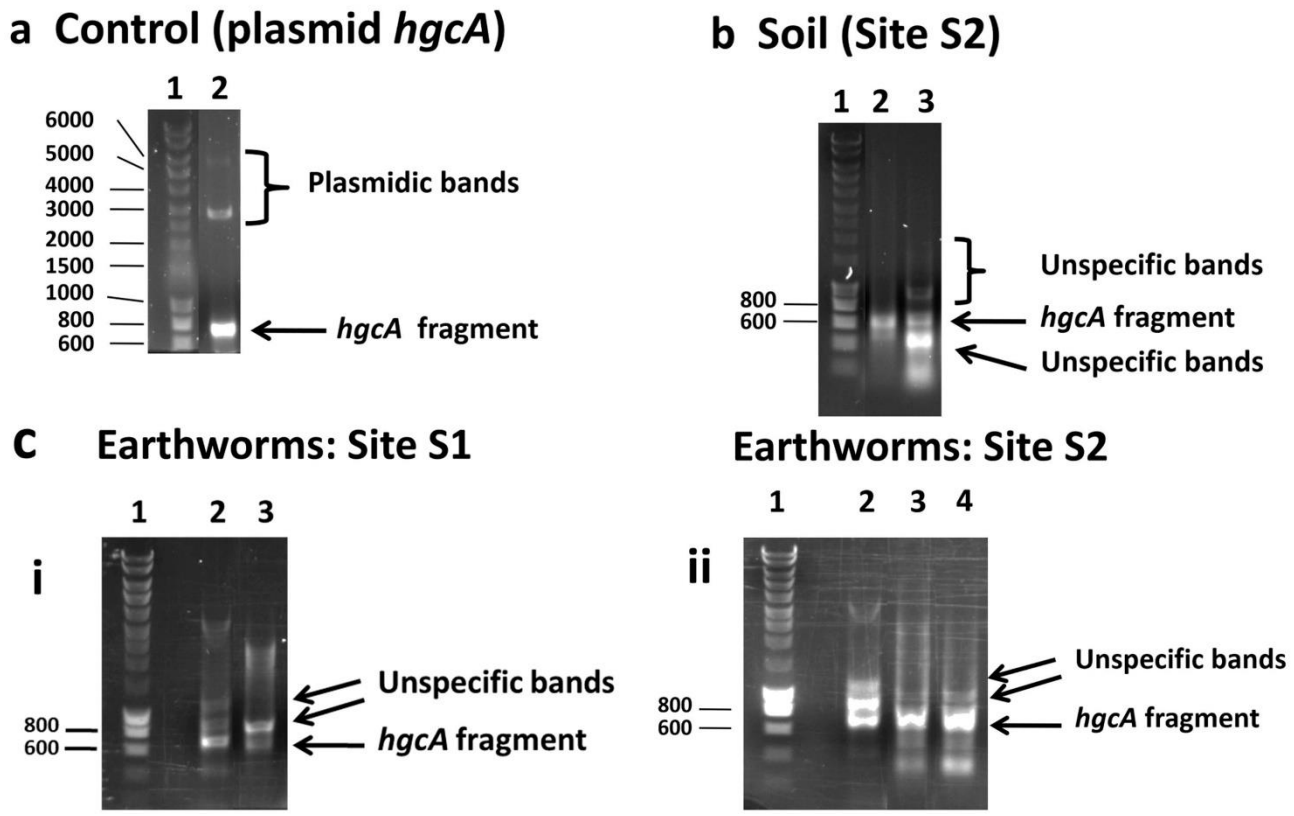
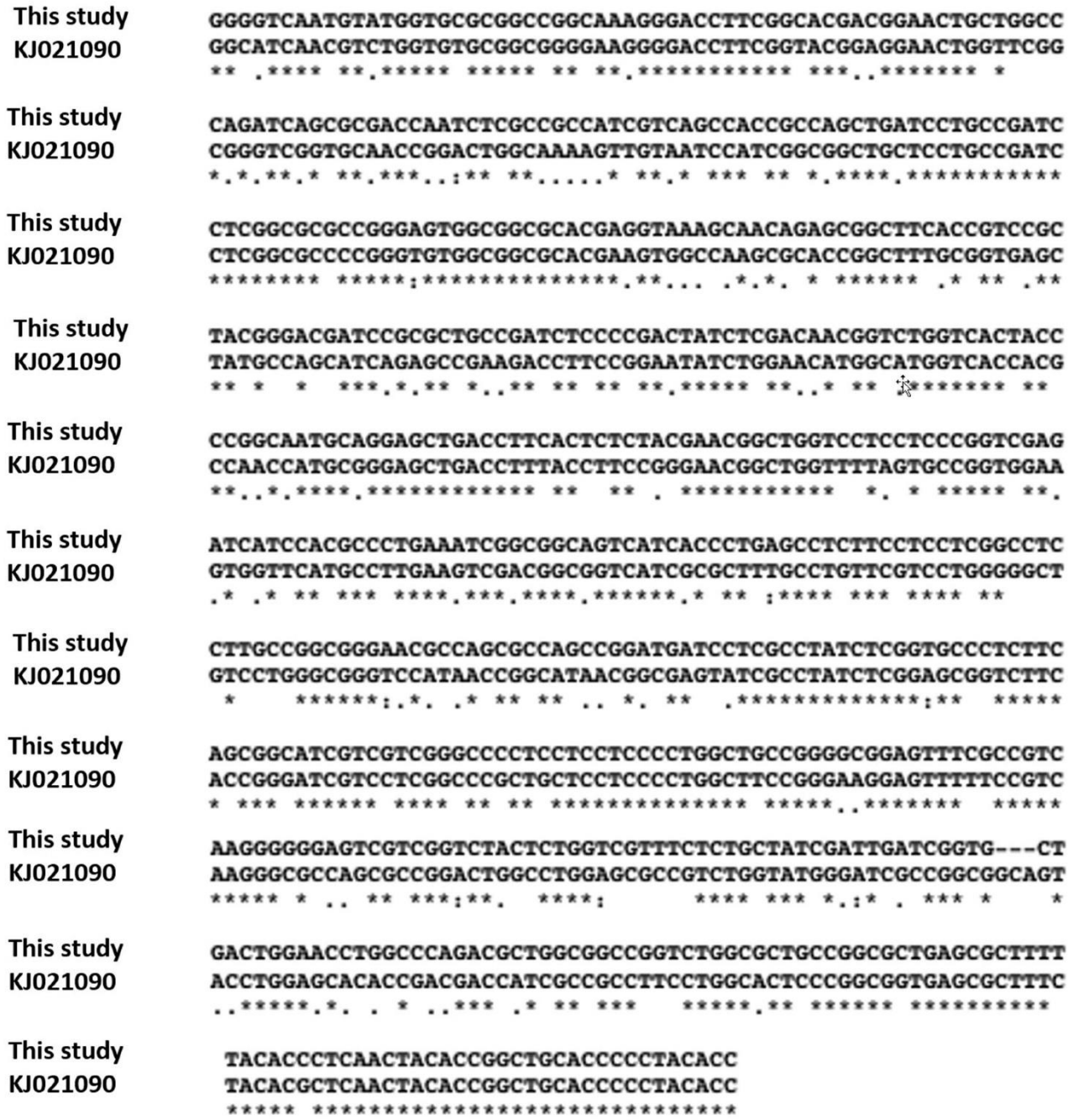


Figure S11. Agarose gel electrophoresis results. a) Shows the control amplification of an *hgcA* gene fragment of 650 bp from a plasmid harboring a cloned fragment of the *hgcA* gene using primers hgcA-F and hgcA-912R and DreamTaq polymerase. Lane 1 – molecular ladder; lane 2 – PCR made up with 200 ng of plasmid and primers hgcA-F + hgcA-912R; band at approximately 650-700 bp. b) Shows the amplification of an *hgcA* gene fragment of approximately 630 bp from total genomic DNA extracted from soil at site S2. Lane 1 – molecular ladder; lane 2 – control PCR with *hgcA*-harboring plasmid (40 ng per reaction); lane 3 – PCR made up with total genomic DNA extracted from the site S2 soil (120 ng of genomic DNA per reaction were used). c) Shows the amplification of *hgcA* gene fragments of approximately 630 to 650 bp from earthworm digestive microflora from sites S1 (panel i) and S2 (panel ii). Panel i, Lane 1 – molecular ladder; lane 2 – control PCR with *hgcA*-harboring plasmid (40 ng per reaction); lane 3 - PCR of the total genomic DNA extracted from the microflora of an anecic earthworm from site S1. Panel ii, Lane 1 – molecular ladder; lane 2 – control PCR with *hgcA*-harboring plasmid (40 ng per reaction); lanes 3 and 4 – PCRs of the total genomic DNA extracted from the microflora of earthworms from site S2, either anecic (lane 3) or endogeic (lane 4).

a. Nucleotide (633 bp) and translated protein sequences (211 amino acids) of the *hgcA* gene fragment amplified from soil and earthworm DNA

```
ggggtcaatgtatggtgcgcggccggcaaagggaccttcggcacgacggaactgctggcc
 G  V  N  V  W  C  A  A  G  K  G  T  F  G  T  T  E  L  L  A
cagatcagcgcgaccaatctcgccgccatcgtcagccaccgccagctgatcctgccgatc
 Q  I  S  A  T  N  L  A  A  I  V  S  H  R  Q  L  I  L  P  I
ctcggcgcgccgggagtggcggcgcacgaggtaaagcaacagagcggcttcaccgtccgc
 L  G  A  P  G  V  A  A  H  E  V  K  Q  Q  S  G  F  T  V  R
tacgggacgatccgcgctgccgatctccccgactatctcgacaacggtctggtcactacc
 Y  G  T  I  R  A  A  D  L  P  D  Y  L  D  N  G  L  V  T  T
ccggcaatgcaggagctgaccttcactctctacgaacggctggtcctcctcccggtcgag
 P  A  M  Q  E  L  T  F  T  L  Y  E  R  L  V  L  L  P  V  E
atcatccacgccctgaaatcggcggcagtcatcaccctgagcctcttcctcctcggcctc
 I  I  H  A  L  K  S  A  A  V  I  T  L  S  L  F  L  L  G  L
cttgccggcgggaacgccagcgccagccggatgatcctcgcctatctcggtgccctcttc
 L  A  G  G  N  A  S  A  S  R  M  I  L  A  Y  L  G  A  L  F
agcggcatcgtcgtcgggcccctcctcctcccctggctgccggggcggagtttcgccgtc
 S  G  I  V  V  G  P  L  L  L  P  W  L  P  G  R  S  F  A  V
aaggggggagtcgtcggtctactctggtcgtttctctgctatcgattgatcggtgctgac
 K  G  G  V  V  G  L  L  W  S  F  L  C  Y  R  L  I  G  A  D
tggaacctggcccagacgctggcggccggtctggcgctgccggcgctgagcgcttttttac
 W  N  L  A  Q  T  L  A  A  G  L  A  L  P  A  L  S  A  F  Y
accctcaactacaccggctgcacccctacacc
 T  L  N  Y  T  G  C  T  P  Y  T
```

b. Comparison of nucleotide sequences of the *hgcA* gene fragment amplified from soil and earthworm DNA (633 bp) and the best match from the NCBI database (KJ021090, *Geobacteraceae* clone B2_FE06).



c. Comparison of amino acid sequences from the *hgcA* gene fragment amplified from soil and earthworm (211 residues) and the best match from the NCBI database (AHL38020, *Geobacteraceae*).

```
This study      GVNVWCAAGKGTFGTTELLAQISATNLAAIVSHRQLILPILGAPGVAAHEVKQQSGFTVR
AHL38020        GVNVWCAAGKGTFGTGELVRQIAASGLAQVVSHRRLLLPILGAPGVAAHEVARRSGFKVS
                *************** **: **:*:.** :****:*:************** ::***.*

This study      YGTIRAADLPDYLDNGLVTTPAMQELTFTLYERLVLLPVEIIHALKSAAVITLSLFLLGL
AHL38020        YAAIRAADLPRYLDNGMVTVPGMQQLSFTLYERLVLIPVELVLALKSISLACFALFVLFA
                *.:******* *****:**.*.**:*:*********:***:: **** ::  ::**:*

This study      LAGGNASASRMILAYLGALFSGIVVGPLLLPWLPGRSFAVKGGVVGLLWSFLCYRLIG-A
AHL38020        VAGGPAAGITAALAWLGACLTGIVIGPLLLPWLPGRSFAVKGAVAGLAWSVLFYLLTGGD
                :*** *:.    **:*** ::***:******************.*.** **.* * * *

This study      DWNLAQTLAAGLALPALSAFYTLNYTGCTPYT--------
AHL38020        AWSAWLTVAALLALPAVSAFYTLNFTGCTPFTSISGVKKE
                 *.   *:** *****:*******:*****:*
```

Figure S12. a) The nucleotide sequence (633 bp) and translated single-letter amino acid sequence (211 residues) of the amplified *hgcA* gene fragment from soil and earthworms. b) A comparison between the nucleotide sequence of the amplified *hgcA* gene fragment from soil and earthworms and the best matching sequence (NCBI web site, KJ021090, *Geobacteraceae* clone B2_FE06). c) A comparison between the amino acid sequence of the amplified *hgcA* gene fragment from soil and earthworms and the best matching sequence (NCBI web site, AHL38020, *Geobacteraceae*). The asterisks, colon, and period denote residues that are identical, have strongly similar properties, and have weakly similar properties, respectively.

a Multiple sequence alignment (various unrelated species)

```
sapiens         MWRSLGLALALCLLPSGGTESQDQSSLCKQPPAWSIRDQDPMLNSNGSVTVVALLQASUY
mouse           MWRSLGLALALCLLPYGGAESQGQSSACYKAPEWYIGDQNPMLNSEGKVTVVALLQASUY
grebe           MWAGLGLVLALCLLPGGGTEIQN----CKEPPEWHIGEENPMLSSRGSVTVVALLQASUY
zebrafish       MWKALSLTLALCLLVGCSAESETEGARCKLPPEWKVGDVEPMKNALGQVTVVAYLQASUL
                ** .*.*.******   .:* :     *  .* * : : :** .: *.***** *****

sapiens         LCILQASKLEDLRVKLKKEGYSNISYIVVNHQGISSRLKYTHLKNKVSEHIPVYQQEENQ
mouse           LCLLQASRLEDLRIKLESQGYFNISYIVVNHQGSPSQLKHSHLKKQVSEHIAVYRQEEDG
grebe           MCLLQASRLEDLRVKLENEGLVNISYVVVNHQGSYSQRKFHLLKERVSNYITVYQQDEQQ
zebrafish       FCLEQASKLNDLLLKLEKQGYPNIAYMVVNNREERSQRLHHLLQERLLN-ITLYAQDLSQ
                :*: ***:*:** :**:.:*  **:*:***::   *:  .  *:::: : *.:* *: .

sapiens         TDVWTLLNGSKDDFLIYDRCGRLVYHLGLPFSFLTFPYVEEAIKIAYCEKKCGNCSLTTL
mouse           IDVWTLLNGNKDDFLIYDRCGRLVYHLGLPYSFLTFPYVEEAIKIAYCEERCGNCNLTSL
grebe           ADVWTTLNGNKDDFLIYR-CGRLVYHLGLPYSFLSFQYVEESIKIAYCENKCGNCSYTEP
zebrafish       PDAWQAVNAEKDDILVYDRCGRLTYHLSLPYTILIHPHVEEAIKHTYCDRICGECSLESS
                 *.*  :*..***:*:*  ****.***.**:::* . :***:** :**:. **:*.

sapiens         KDEDFCERVSLATVDKTVET--------PSPHYHHEHHHNHGHQHLGSSELSENQQPGAP
mouse           EDEDFCKTVTSATANKTAEP--------SEAHSHHKHHNKHGQEHLGSSKPSENQQPGPS
grebe           AIDDICENITKTADEKLAEPEPKPTGQHSHHHNLHRHRHHHHHHHREGSRHSKNENHQNS
zebrafish       AQLEECKKATEEVNKPVEEE----------PRQDHGHHEQGHHEHQGEAERHRHG-----
                   : *:  :  . .   *            :  * *:.:  :.*   :.  .:

sapiens         NAPTHPAPPGLHHHHKHKGQHRQGHPEN--RDMPASEDLQDLQKKLCRKRCINQLLCKLP
mouse           --ETTLPPSGLHHHHRHRGQHRQGHLESUDTTASEGLHLSLAQRKLURRGCINQLLCKLS
grebe           SETQRHHPHSGHNRHDHTGSHEQVDTVPPGESVEISQDKKLU--KKGKTSCKNQLTUNWQ
zebrafish       ----HHHPHHHHHHHRGQQQVDVDQQVLSQVDFGQVAVETPMMKRPUAKHSRUKVQYSUQ
                       *   *::*    .    .                    :     .  ::  .

sapiens         TDSELAPRSUCCHCRHLIFEK-TGSAIT-UQCKENLPSLCSUQGLRAEENITES-CQURL
mouse           KESEAAPSSCCCHCRHLIFEK-SGSAIA-UQCAENLPSLCSUQGLFAEEKVTES-CQCRS
grebe           TASDSASSSUCCHURHLLFEE-LGNSVT-UQCRGALPNSCRUHGQLSAEDITES-UQURL
zebrafish       QGADSPVASUCUHURQLFGGEGNGRVAGLUHCDEPLPASUPUQGLKEQDNHIKETUQURP
                  :: .  *:*:* *:*:  :  *     *:*   **  : *:*    :.  :. :*:*

sapiens         -PPAAUQISQQLIPTEASASURUKNQAKKUEUPSN
mouse           -PPAAUQ-NQPMNPMEANPNUSUDNQTRKUKUHSN
grebe           -LTAAUQ-SPAAGASETSDTUQUQEKARNUAUKTN
zebrafish       APPAEU--------ELSQPTUVUPAGDATUGURKK
                 ..* *: .           * *      * *
```

b **Multiple sequence alignment (various bird species)**

```
grebe          MWAGLGLVLALCLLPGGGTEIQNCKEPPEWHIGEENPMLSSRGSVTVVALLQASUYMCLL
crane          MWAGLGLVLALCLLPGGGTEIQNCKEPPEWHIGEENPMLNTRGSVTVVALLQASUYLCLL
pelican        MWAGLGLVLALCLLPGGGTEIQNCKEPPEWHIGEENPMLNSRGSVTVVALLQASUYLCLL
hoatzin        MWAGLGIVLALCLLPGGGTEIQNCKEAPEWRIGEEDPMLNSRGSVTVVALLQASUYLCVL
penguin        MWAGLGLVLALCLLSGGGTEIQNCKEPPEWRIGEENPMLNSRGSVTVVALLQASUYLCLL
emperor        MWAGLGLVLALCLLSGAGTEIQNCKEPPEWRIGEENPMLNSRGSVTVVALLQASUYLCLL
cormorant      MWAGLGLVLALCLLPGGGTEIQNCKEPPEWRIGEENPMLNSRGSVTVVALLQASUYLCLL
ibis           MWAGLGLVLALCLLPGGGTEIQNCKEPPEWRIGEENPMLNSRGSVTVVAPLQASUYLCLL
tropicbird     MWAGLGLVLAVCLLPGGGTEIQNCREPPEWRIGEENPMLNSRGSVTVVALLQASUYLCLL
loon           MWAGLGLVLALCLLPGGGTEIQNCKEPPEWRIGEENPMLNSRGSVTVVALLQASUYLCLL
fulmar         ------------------------------------------------------------
               ******:***:***.*.*******:*.***:****:***.:******** ******:*:*

grebe          QASRLEDLRVKLENEGLVNISYVVVNHQGSYSQRKFHLLKERVSNYITVYQQDEQQADVW
crane          QASRLEDLRVKLENKGLFNISYVVVNHQGTYSQKKFHLLKERVSEYITVYQQDEQQDDVW
pelican        QASRLEDLRVKLQNEGLVNISYVVVNHQGTYSQRKFHLLKESVSDYITVYQQDEQQADVW
hoatzin        QASRLEDLRVKLVNEGLVNISYVVVNHQGPSSRRKFHLLKESVSDSITVYQQDEQQDDVW
penguin        QASRLEDLRVKLENEGLVNISYVVVNHQGTYSQRKFHLLKESVSDYITVYQQDEQQADVW
emperor        QASRLEDLRVKLENEGLVNISYVVVNHQGTYSQRKFHLLKESVSDYITVYQQDEQQADVW
cormorant      QASRLEDLRVKLENEGLVNISYVVVNHQGTSSQRKIHLLKASVSDYVTVYQQDEQQADVW
ibis           QASRLEELRVKLENEGLVNISYVVVNHQGTYSQRKFHLLKESVSDYITVYQQDEQQADVW
tropicbird     QASRLEDLRVKLQNEGLVNISYVVVNHQGTSSQRKFHLLKESVSDYITVYQQDEQQADVW
loon           QASRLEDLRVKLENEGLVNISYVVVNHQGTSSQRKFHLLKESVSDSITVYQQDEQQADVW
fulmar         ------------------------------------------------------------M
               ******:***** *:**.***********. *:::*:****  **: :********* **

grebe          TTLNGNKDDFLIYR-CGRLVYHLGLPYSFLSFQYVEESIKIAYCENKCGNCSYTEPAIDD
crane          TTLNGNKDDFLIYDRCGRLVYHLGLPYSFLSFQYVEESIKIAYCENKCGNCSYTEPDIDV
pelican        TTLNGNKDDFLIYDRCGRLVYHLGLPYSFLSFQYVEESIKIAYCENKCGNCSYTEPDIDG
hoatzin        TTLNGNKDDFLIYDRCGRLVYHLGLPYSFLSFQYVEDSIKIAYCENKCGNCSYTEPGIDG
penguin        TTLNGNKDDFLIYDRCGRLVYHLGLPYSFLSFQYVEESIKIAYCENKCGNCSYMEPEIDG
emperor        TTLNGNKDDFLIYDRCGRLVYHLGLPYSFLSFQYVEESIKIAYCENKCGNCSYMEPEIDG
cormorant      TTLNGSKDDFLIYDRCGRLVYHLGLPYSYLSFQYVEESIKIAYCENKCGNCSYMEPDIDG
ibis           TTLNGNKDDFLIYDRCGRLVYHLGLPYSFLSFQYVEESIKIAYCENKCGNCSYTEPDIDG
tropicbird     TTLNGNKDDFLIYDRCGRLVYHLGLPYSFLSFQYVAESIKIAYCENKCGNCSYMEPDIDG
loon           ATLNGNKDDFLIYDRCGRLVYHLGLPYSFLSFQYVEESIKIAYCENKCGNCSYMEHDIDG
fulmar         GQIGTIKIPSRTLDRCGRLVYHLGLPYSFLSFQYVEESIKIAYCENKCGNCFYMEPDIGG
                 :.  *        *************:****** :************** * *  *.

grebe          ICENITKTADEKLAEPEPKPTGQHSHHHNLHRHRHHHHHHHHREGSRHSKNENHQNSSETQ
crane          ICENITKKTDEKLSEIEPKPTGQHSHHHSLHRHRHHHHHHHR-EGGRHSKNENHQTPSESQ
pelican        VCKNITQKADEKLTEIEPKPTGQHSHHHSLHRHRQHHHREG---SRHSKNENHQASSETQ
hoatzin        ACENITKKADEKLAEIEPIPTGQHSHHHNPHRHRHHHRHHH-EGSHHSKNENHRAPSETQ
penguin        ICENMTKKADEKLAEVEPKPTGQLSHHHNLHRHRQHHHHRE--GSHHAKNENRQAPSETQ
emperor        LCENITKKADEKLAEIEPKPTGQHSHHHNLHRHRQHHHHRE--GSHHSKNENHQAPSETQ
cormorant      ICENITKQADEKRAEIEPSPTGQRSHDHNTHRHIKHHQHRE--GSHHSKNENHQAPSETR
ibis           ICENITKKADEKLAERELQPADQHSHHHDPHRHGHHHHHRE--GSHHSRNENHQALSETQ
tropicbird     ICENITKKADEKLAEVEPKPTGQHSHHHSLHRHRHHHHHHE--SSRHSKNENHQSPSETQ
loon           VCENITKKADEKLAEIEPTPTGQHSHHHNLHRHRQHHHHRE--SSRHSKNENHQAPSETQ
fulmar         ICENITKKAEEKLAEIEPKPTGQHSHHHNLHRHRQHHNHRE--GSRHSKNENHQAPSETQ
                *:*:*: ::** :* *  *:.* **.*. *** :**.:.    .:*::***::  **::
```

```
grebe           RHHPHS-------GHNRHDHTGS-HEQVDTVPPGESVEISQDKKLUKKGKTSCKNQLTUN
crane           RHHPHSGQHRRVFGHNRHDHTGS-HEQVDTAPPGESVEITQDKKLUKKGKNSCKNQLTUN
pelican         RHHPHSGQRRRVFGRNRHDQTGN-HEQVDTVPAGESVEITQDKKLUKKGKTSCKNQLTUN
hoatzin         RHHPRSGQRHRVSGHNRHGQTGR-DAQVDTVPPGESVEITQDKKLUKKGKTSCKNQLSUN
penguin         RHHPHSGRRHRVFGHNRRDQTDDSHEQVDTVPPGESVEITQDKKLUKKGKTSCKNQLTUN
emperor         RHHPHSGRRHRVFGHSRRDQADS-QEQVDSVPPGESVEITQDKKLUKKGKTSCKNQLTUN
cormorant       RHHSDSSQRHRAFGRNRHVQTGS-HEQVDTVPPGESVEITQDKKLUKKGKISCKNQLTUN
ibis            RRHAHSSRRHRVFGRNRHDQTGS-HEQVDTVPPGESVDITQDKKLUKKGKTSCKNQLTUN
tropicbird      RHHPHSSRRHRVFGHNTHDQTGS-HEQVDIGPPGESVEMTQDKKLUKKGKNSCKNQLTUH
loon            RHHPHSGRRHRVLGRTRHDQTGS-HEHVDTVPPGESVEITQDKKLUKKGKTSCKKQLTUN
fulmar          RHHPHSGRRHRVFGRNRHDQTGS-HEQVDTVPPGESVEITQDKKLUKKGKTSCKNQLTUS
                *:*. *       *:. : ::.  . :**  *.****:::********** ***:**:*

grebe           WQTASDSASSSUCCHURHLLFEELGNSVTUQCRGALPNSCRUHGQLSAEDITESUQURLL
crane           WQTASDSTSSSUCCHCRHLLFEELGNSVTUQCRGALPNSCRUHGQLSAEDITESUQURLL
pelican         WQTASDSTSSSUCCHURHLLFEELGNSVTUQCRGALPNSCRUHGQLPAEDITESUQURLL
hoatzin         WQTASDSNSSSUCCHCRHLLFEELGNSVTUQCRGALPNSCRUHGQLAAEDITESUQCRLL
penguin         WQTASDPTSSSUCCHCRHLLFEELGNSVTUQCHGALPNSCSUHGQLSAEDITESUQURLL
emperor         WQTASDSTSSSUCCHCRHLLFEELGNSVTUWCHGALPNSCSUHGQLSAEDITESUQURLL
cormorant       WQTASDSTSSSUCCHCRHLLFEELGNSVTUQCRGALPNSURUHGQLSAEDITESUQURLL
ibis            WQTASDSTSSSUCCHCRHLLFEELGNSVTUQCRGALPNSCRUHGQLSAEDITESUQURLL
tropicbird      WQTASDSASSNUCCHCRHLLFEELGNSVTUQCRGALPSSCRUHGQLSAEDITESUQURLL
loon            WQTASDLTSSSUCCHCRHLLFEELGNSVTUQCRGALPNSCRUHGQLSAEDITESUQURLL
fulmar          WQTASDSASSSUCCHCRHLLFEELGNSVTUQCRGALPNSCRUHGQLSAEDITESUQURLL
                ******  **.****:*************** *:****.*: *****.*********:***

grebe           TAAUQSPAAGASETSDTUQUQEKARNUAUKTN
crane           TAAUQSPAAGTSEASDTUQUQEKARNUAUKTN
pelican         TAAUQSPAA--SETSDTUEUQEKARNUAUKTN
hoatzin         TAAUQSPAAGGSETSDTUQUQENARNUAUKTN
penguin         TAAUQSPAA--SETRDTUQUQEKARNUAUKTN
emperor         TAAURSPAA--SETSDTUQUQEKARNUAUKTN
cormorant       TAAUQSPA--TSETSDTUHUQEKARNUAUKTN
ibis            TAAUQSPAAGVSDSSDTUQUQEKARNUAUKTN
tropicbird      TAAUQSPAAGATETSDTUQUQEKARNUAUKTN
loon            TAAUQSPAAGASETSDTUQUQEKARNUAUKTN
fulmar          TD-UQSPTAGASETSDTUQUQEKARNUAUKTN
                *  *:**::: :  *** ***:*********
```

C **Multiple sequence alignment (fish species)**

```
amberjack       -MWACLSLLLTLCLLHGGGAESDGDGPRCQPPPAWKIGEVDPMKEALGQVTVVALLQASU
eel             -MWACLSLLLPLCLLHGGRAESEGVGPRCQLPPAWNIGEVEPMKGTMGQVTVVALLQASC
perch           -MWRGLSLLLTLCLLHGGGAESDGGGPRCQLPPAWKIGEVEPMTQAMGRVTVLALFQASU
tilapia         -MWAGLSLLLTLCLLHGGGAESEGGGPRCQLPSDWRIGDVEPMKGSVGRVTVVALLQASU
pike            -MWAGLSLLLALCLLPGGGTESEGEGSRCKPPPIWSIGEVEPMKEAMGQVTVVALLQASU
trout           MMWVGLSLLLALCLLPGGGTESEGEGTRCKQPPGWSIGEVEPMKEVMGQVTVVALLQASU
salmon          -MKAGLSLLLALCLLPGGGAESEGEGTRCKPPAGWSIGEVEPMKGVMGQVTVVALLQASU
zebrafish       -MWKALSLTLALCLLVGCSAESETEGARCKLPPEWKVGDVEPMKNALGQVTVVAYLQASU
                 *   *** :.**** *  :**:  *.**: *. * :*:*:**.  :*:***:* :***:

amberjack       LFCLVQASRVDGLRQRLESQGLKDVTYMVINHQGEEAQRLHAVLAQRLSENITLYKQEEQ
eel             LFCLVQASRMDSLHQMMESQGLKNVAYMVINHQGEQAQLLHPMLAQRMSENITLYKQDQQ
perch           LFCLVQASRMDPLRNKLESQGLKNVVYMVVNHQGEHAQRLHPMLADRLSENITLYKQDEL
tilapia         LFCLVQASRLDGLQQKLERQGLKNVVYMVVNHQGEQSRHLHPLLEAKLSKNIILYKQDGH
pike            LFCLVQASLLDGLRLKLEGQGLKNVHYMVVNHQGEKAQRLHKLLKQKLSEKITLYKQEPE
trout           LFCLVQASLLDGLRLKLEGQGLENVTYMVVNHQGEQAQRLHTLLRQKLSENITLYKQQPK
salmon          LFCLVQASLLDELRLKLEGQGLDNVTYMVVNHQGDQAQHLHTLLSQKLSENIILYKQEPK
zebrafish       LFCLEQASKLNDLLLKLEKQGYPNIAYMVVNNREERSQRLHHLLQERLLN-ITLYAQDLS
                **** **. :: *   :* **  :: ***:*::  .:: ** :*  :: : * ** *:

amberjack       QPDVWQTLSGEKDDFLIYDRCGRLTHHISLPYSIIGQGHVEGAIKDTYCKRICGNCTHES
eel             QVDVWKTLGGQKDDFLIYDRCGRLTHHISLPYSIIGQGHVEGAIRDTYCNRICGDCTHES
perch           QPDVWKSLNAQKDDFLIYDRCGRLTHHISLPYSIIGQGHVESAIKEAYCKRMCGDCTHES
tilapia         QPDVWQTLAGEKDDFFIYDRCGRLTYRISLPYSIIGEGHIEKAIKDTYCKRLCGDCTHES
pike            QVDVWQSLAGQKDDFLIYDRCGRLTYHISLPYSIMSIPYVENAIKETYCARICGNCKHES
trout           QEDVWQTLAGEKDDFLIYDRCGRLTYHISLPYSILGTPYVENAIKETYCTRVCGDCTYES
salmon          RADVWQALAGKKDDFLIYDRCGRLTHHIFLPFSILGTPYVENAIKETYCQSICGDCTYES
zebrafish       QPDAWQAVNAEKDDILVYDRCGRLTYHLSLPYTILIHPHVEEAIKHTYCDRICGECSLES
                : *.*::: .:***:::******** :: **::*:   ::* **:.:**  :**:*. **

amberjack       TDTPEECKGKADAQPTTDGTPAVEEDTGHGHGHGHGHGHGHGLG---HGHHHG-------
eel             TETPEECKGKADVQPDADGTPAIEHNTGHGHHHGHHHGHGHGHHGDNSGFHHG-------
perch           ADIPEECRETADAQPGADSPPAVEEGTQHGHSQDHAHGHHHRHHHHHGHGHQGDNNDFQS
tilapia         AEIPEECKDNAGVQPDVPAEQDDTR-HDHHHGHGHGHHHGHGHHHGHGHGHHGDNQDVHP
pike            VEIPAECNGTAEATSEGEDKPTTTVEPTHDGHQHHRHHHHHH-------HHDDKHGDHGD
trout           KEIPAECNRTVEAKPEGEEKPVTGRETTHGGHGHHHHGHGHNGNRHGHNGNRHGHDHHGE
salmon          TEIPAECNRMVEVKPEGEEKPVTGGDTPHGGRGHHHHGNGHG-----HHSKSHGHGHHGE
zebrafish       SAQLEECKKATEEVNKPVEEEPRQDHGHHEQGHHEHQGEAERHR----------------
                     **.  .                 :     . : . .
```

```
amberjack       --HHHGHN--HGFHPRGFGHGRDHGTQGQEQ--DSGVSQGQQQAELGQMQQAVQVHLVSQ
eel             --HGHDHN--HGHHHGGDDVAGQG--HDHRQ--VGGVSQRQEHFDLGQMQQAVLTQQMLQ
perch            RGHGHHHGHHHGHHHQHHHGQSQQGVSDQGHEQDGGLSQRQPSLDLGQMQQEVHLQEMSH
tilapia          HGHGSDHNNGHHHRNHDGADQTQHGVRPHGHFHEGDMPQTQHHFDLGQIPQEVHNQQVAQ
pike            REVGRGHGAEQ-----QRHHKHVGEGHRHVQD-----QLHVSQDHVGQTAVQLG---QET
trout           RGMGRGHGRDHGAER-QHQHDTEGLQHGQAHG-----QLHVGQEHMGQQAVQLGQMPQEG
salmon          SEVGRDHGRGHGVEQQQHQHGAEGLHHGQAHG-----QLHVGQEHMGQQ-------PKEA
zebrafish       --HGHHHPHHHHHHHRGQQQ----------------------VDVDQQVLSQVDFGQVA
                      *                                    .*

amberjack       EANGAP-VRPUVPQKARUKSKYSUQWTAGSDNEASPKVSUCUHURRLFGEAGSKQPVGLU
eel             EAHRAP-VRPUASERARUKSKHSUQWTAGSDNEASPKISUCUHURRLFGDAGSEQPVGLU
perch           QAQAEP-SRPUVQETVRCKSKHSUQWAAGSDNEASPKASUCUHURKLFDDAGSEQPVGLU
tilapia         EAHAVI-ERPULSRKNRUKLKYNUQGLTGSDNEI--KSSUCUHURRLFGEAGSEQPVGLU
pike            NEGQVM-QRPUAKEGVRCSGQHDUQWKEGSVLSPSSKASUCUHURHLLDDGVSGQPIRLU
trout           QRGHIM-QNPUVKGKSRUKAEHSUQWKEGSDLSPSSKASUCUHURRLFGDGVSNEPIGLU
salmon          QEGHIM-QRPUVKGRARUKAELSUHLKEGSDISPSSKVSUCUHURGLFGNGVSNEPVGLU
zebrafish       VETPMM-KRPUAKH-SRUKVQYSUQQ--GAD---SPVASUCUHURQLFGGEGNGRVAGLU
                       :**     *:   : **   *     :   ******* *:   . .    **

amberjack       HCDKALPASUQUHGLIGDAANNVRETUQURSPPAE-UQQPQPAQUAUPPGVVS------
eel             HCNEALPASUQUHGLIGDAASNVRETUQURTPPAD-UQQPQPNQUAUPPGVVS------
perch           HCDEALPASUQUHGLKGDAA--VRETUQURSPPAD-UQQPQPAQUAUPPGVVS------
tilapia         HCDEALPTSURUHGLIGDAVNDVRETUQURLPHAA-UQEPQPAQUAUPPGVVS------
pike            HCNEALPASCQUHGLKG----LLRETUQURSPQTD-UQQPRPVMUAUSPGVES------
trout           HCDEALPASUQUQGLIG-----LRETUQURSSLAD-UQQPQPVMUAUPLGVESUGUGLL
salmon          HCDEALPASUQUQGLMGDSTNHIRETUQURSPPAD-UQQPPPVMSAUSPGVETUGUEQL
```

d Multiple sequence alignment (various cetacean species)

```
dolphin         MNIRIIAPAMWRGLGLALALCLLPTGGTESQGQSSFCKQPPAWSIKDQDPMLNSYGSVTV
tursiops        MNIRMTAPAMWRGLGLALALCLLPTGGTESQGQSSFCKQPPAWSIKDQDPMLNSYGSVTV
orca            ---------MWRGLGLALALCLLPTGGTESQGQSSFCKQPPAWSIKDQDPMLNSYGSVTV
beluga          MNIRMTAPAMWRGLGLALALCLLPTGGTESQGQSSFCKQPPAWSIKDQDPMLNSYGSVTV
porpoise        ---------MWRGLGLALALCLLPTGGTESQGQSSFCKQPPAWSIKDQDPMLNSYGSVTV
cachalot        ----MTAPAMWRGLGLALALCLLPTGGTESQGQSSFCKQPPAWSIKDQDPMLNSYGSVTV
lipotes         ---------MWRGLGLALALCLLPTGGTESQGQSSFCKQPPAWSIKDQDPMLNSYGSVTV
                         **************************************************

dolphin         VALLQASUYLCILQASRLEDLRVKLEKEGYSNISYVVVNHQGISSRLSYVHLKNKVSEHI
tursiops        VALLQASUYLCILQASRLEDLRVKLEKEGYSNISYVVVNHQGISSRLSYVHLKNKVSEHI
orca            VALLQASUYLCILQASRLEDLRVKLEKEGYSNISYVVVNHQGISSRLSYVHLKNKVSEHI
beluga          VALLQASUYLCILQASRLEDLRVKLEKEGYSNISYVVVNHQGISSRLNYVHLKNKVSEHI
porpoise        VALLQASUYLCILQASRLEDLRVKLEKEGYSNISYVVVNHQGISSRLNYVHLKNKVSEHI
cachalot        VALLQASUYLCILQASRLEDLRVKLEKEGYSNISYVVVNHQGISSRLKYVHLKNKVSEHI
lipotes         VALLQASUYLCILQASRLEDLRVKLEKEGYSNISYVVVNHQGISSRLKYVHLKNKVSEHI
                **********************************************.************

dolphin         PVYQQEENQTDVWTLLNGNKDDFLIYDRCGRLVYHLGLPYSFLTFSYVEDSIKSVYCESK
tursiops        PVYQQEENQTDVWTLLNGNKDDFLIYDRCGRLVYHLGLPYSFLTFSYVEDSIKSVYCESK
orca            PVYQQEENQTDVWTLLNGNKDDFLIYDRCGRLVYHLGLPYSFLTFSYVEDSIKSVYCESK
beluga          PVYQQEENQTDVWTLLNGKKDDFLIYDRCGRLVYHLGLPYSFLTFSYVEDSIKTVYCESK
porpoise        PVYQQEENQTDVWTLLNGKKDDFLIYDRCGRLVYHLGLPYSFLTFSYVEDSIKTVYCESK
cachalot        PVYQQEENQTDVWTLLNGKKDDFLIYDRCGRLVYHLGLPYSFLNFSYVEDSIKTVYCESK
lipotes         PVYQQEENQTDVWTLLNGKKDDFLIYDRCGRLVYHLGLPYSFLTFSYVEDSIKTVYCESK
                ******************:************************.*********:******

dolphin         CGNCSLKTLEYEDFCKNVSLATVEKTTEAPQP------HHHHHHHHHLENGHLSEYPKPE
tursiops        CGNCSLKTLEYEDFCKNVSLATVEKTTEAPQP------HHHHHHHHHLENGHLSEYPKPE
orca            CGNCSLKTLEYEDFCKNVSLATVEKTTEAPQPHHHHHHHHHHHHHHHHLENGHLSEYPKPE
beluga          CGNCSLKTLEDEDFCKNVSLATAEKTTEAPQP---HHHHHHR-----LENSHLSEYPKPE
porpoise        CGNCSLKTLEDEDFCKNVSLATAEKTTEAPQP---HHHHHHHHHHHHLENSHLSEYPKPE
cachalot        CGNCSLKTLEDEGFCKNVSLATEEKTTEAPQP--HHHHHHHHHHHHHLENGHLSEYPIPE
lipotes         CGNCSLKTLEDEGFCKDVSLATAEKTTEAPQP--HHHHHHHHHHHHHLENGHLSEYPKPE
                ********** *.***:***** *********      ***:     ***.****** **
```

```
dolphin          APDNPEHPPPSGLHRHHRHRGHQRQGHSENUDALQLSFPQKKLURSGCINQLLUKFPKDS
tursiops         APDNPEHPPPSGLHRHHRHRGHQRQGHSENUDALQLSFPQKKLURSGCINQLLUKFPKDS
orca             APDNPEHPPPSGLHHHHRHRGHQRQGHSENUDALQLSFPQKKLURSGCINQLLUKFPKDS
beluga           APDNPEHPPPSGLHRHHRHRGHQRQGHSENUDALQLSLPQKKLURSGCINQLLUKFPKDS
porpoise         APDNPEHAPPSGLHRHHRHRGHQRQGHSENUDALQLSLPQKKLURSGCINQLLUKFPKDS
cachalot         APDNPEHTPPSGLHRDHRHRGHQRQGHSENUDTLQLSLPQKKLUVRGCVNQLLUKFPRDS
lipotes          APDNPEHPPPSGLHHHHRHRGHQRQGHSENUDTLQISLPQKKLURSGCINQLLUKFPKDS
                 *******.******:.****************:**:*:******  **:********:**

dolphin          ESALSSUUUHCRHLIFEKTGSAVTUQCKDNLPSLCSUQGLLAEEHVTESUQURLPPAACH
tursiops         ESALSSUUUHCRHLIFEKTGSAVTUQCKDNLPSLCSUQGLLAEEHVTESUQURLPPAACH
orca             ESALSSUUUHCRHLIFEKTGSAVTUQCKDNLPSLCSUQGLLAEEHVTESUQURLPPAACH
beluga           ESALSSUCUHCRHLIFEKTGSAVTUQCKQNLPSLCSUQGLLAEEHVTESUQURLPPAACH
porpoise         ESALSSUCUHCRHLIFEKTGSAVTUQCKENLPSLCSUQGLLAEEHVTESUQURLPPAACH
cachalot         ESALSSUUUHCRHLIFEKTGSAVTUQCKENLPSLCSUQGLLAEEHVTESUQURLPRAACP
lipotes          ESALSSUCUHCRHLIFEKTGSAVTUQCKENLPSLCSUQGLLAEEHVTESUQURLPPAACQ
                 ******* ********************:************************** ***

dolphin          SSQQLKPTEAGTKUSUEKKAEMUKUPSD
tursiops         ATQQLKPTEASTKUSUEKKAEMUKUPSD
orca             ASQQLKPTEASTKUSUEKKAEMUKUPSD
beluga           ASQQLKPTEASTKUSUEKKAEMUKUPSN
porpoise         ASQQLKPTEASTKUSUEKKAEMUKUPSN
cachalot         ASQQLNPTEASTKUSUKKKAEMUKUPSN
lipotes          ASQQPNPTEASTKUSUKKKAEMUKUPSN
                 ::** :****.*****:**********:
```

e Multiple sequence alignment (various pinniped species)

```
fur seal         MWRSLGLALALCLLPWGGAESQGQSSLCKPPPAWSIGDQNPMVKSSGSVTVVALLQASUY
sea lion         MWRSLGLALALCLLPWGGAESQGQSSFCKPPPAWSIGDQNPMVNSSGSVTVVALLQASUY
walrus           MWRSLGLALALCLLPWGGAESQGQSSFCKPPPAWSIGDQNPMVNSSGSVTVVALLQASUY
Weddell seal     MWRSLGLALALCLLPWGGAESQEQSSFCKQPPAWSIGDQNPMLNSSGSVTVVSLLQASUY
monk seal        MWRSLGLALALCLLPWGGAESQEQSSFCKQPPAWSIGDQNPMLNSSGSVTVVALLQASUY
                 ********************** ***:** ************::********:*******

fur seal         LCILQASRLEDLRVKLEKEGYLNISYVVVNHQGLSSQLKYMHLKNKVSEYIAVYQQEENQ
sea lion         LCILQASRLEDLRVKLEKEGYLNISYVVVNHQGLSSQLKYMHLKNKVSEYIAVYQQEENQ
walrus           LCILQASRLEDLRIKLEKEGYLNISYVVVNHQGLSSQLKYMHLKNKVSEYIAVYQQEENQ
Weddell seal     LCILQASRLEDLRVKLEKEGYLNISYVVVNHQGLSSQLKYMHLKNKVSEYIPVYQQEENQ
monk seal        LCILQASRLEDLRVKLEKEGYLNISYVVVNHQGLSSQLKYMHLKNKVSEYIPVYQQEENQ
                 *************:************************************.********

fur seal         TDVWMLLNGKKEDFLIYDRCGRLVYHLGLPYSFLTFPYIEEAIKITYCKEKCGSCSLMTL
sea lion         TDVWMLLNGKKEDFLIYDRCGRLVYHLGLPYSFLTFPYIEEAIKITYCKEKCGNCSLMTL
walrus           TDVWTLLNGKKEDFLIYDRCGRLVYHLGLPYSFLTFPYVEEAIKITYCKERCGSCSLTTL
Weddell seal     TDVWMLLNGKKDDFLIYDRCGRLVYHLGLPYSFLTFPYVEEAIKFTYCKEKCGHCSLMTL
monk seal        TDVWMLLNGKKDDFLIYDRCGRLVYHLGLPYSFLTFPYVEEAIKITYCEEKCGHCSLMTL
                 **** ******:****************************:*****:***:*:** *** **

fur seal         ESEDFCKMVSSATVGNTTEAPQP---HHHHHHHHHRHHHKHGHQHHGIDGVSENQQPEAP
sea lion         EAEDFCKMVSLTTVGNTTEAPQPH--HHHHHHHHHCHHHKHGHQHHGIDGVSENQQPEAP
walrus           EAEDFCKMVCLATVGDTTEAPQPHHHHHHHHHHHHHHRHHHKHGHQHHGIDGVSENQQPEAP
Weddell seal     EAADHCKMVSSATVGNTTEAPQP-----HHHHHHHGHHLKHEHQHHGIDGVSENQQPEAP
monk seal        EAEDHCKMVSSATVGNTTEAPQP-----HHHHHHHGHHHKHEHQHHGIDGVSENQQPEAP
                 *: *.****. :***:*******     ******* ** ** ******************
```

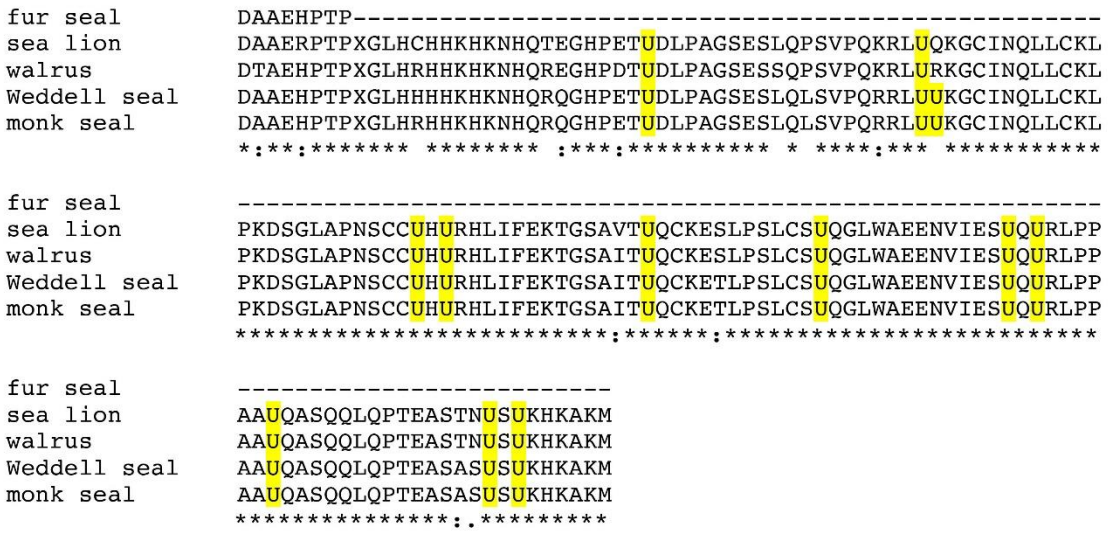


Figure S13. Multiple sequence alignments of SelP proteins. Selenocysteine residues (U) are highlighted in yellow. The conserved XUXUX^{6}UXUX motif is present in cetaceans, birds, and mammals. This motif is also present in some fish species. The UXUX^{6}UX^{7}UXU motif is found in many species and pinnipeds have the following motif: UXUX^{6}UX^{14}UXU. All these motifs can form a tetraselenolate complex with Hg. The asterisks, colon, and period denote residues that are identical, have strongly similar properties, and have weakly similar properties, respectively. GenBank access numbers are in parenthesis.

a) *Homo sapiens* (AAH15875.1); house mouse: *Mus musculus* (NP_001036079.1); great crested grebe: *Podiceps cristatus* (KL268852.1); zebrafish: *Danio rerio* (NP_840082.3).

b) great crested grebe: *Podiceps crist*atus (KL268852.1); grey crowned crane: *Balearica regulorum gibbericeps* (XP_010311294.1); Dalmatian pelican: *Pelecanus crispus* (XP_009484102.1); Adélie penguin: *Pygoscelis adeliae* (XP_009322697.1); emperor penguin: *Aptenodytes forsteri* (XP_009272808.2); great cormorant: *Phalacrocorax carbo* (XP_009511852.1); crested ibis: *Nipponia nippon* (XP_009472359.1); white-tailed tropicbird: *Phaethon lepturus* (XP_010284851.1); red-throated loon: *Gavia stellata* (XP_009818154.1); Arctic fulmar: *Fulmarus glacialis* (XP_009579693.1).

c) greater amberjack: *Seriola dumerili* (XP_022613554.1); swamp eel: *Monopterus albus* (XP_020449352.1); climbing perch: *Anabas testudineus* (XP_026199022.1); Nile tilapia: *Oreochromis niloticus* (XP_013127321.2); northern pike : *Esox lucius* (XP_012992107.1); rainbow trout: *Oncorhynchus mykiss* (CCX35038.1); Atlantic salmon: *Salmo salar* (AXS67852.2); zebrafish: *Danio rerio* (NP_840082.3).

d) Pacific white-sided dolphin: *Lagenorhynchus obliquidens* (XP_026956984.1); common bottlenose dolphin: *Tursiops truncatus* (XP_019785640.1); killer whale: *Orcinus orca* (XP_004266005.2); beluga whale: *Delphinapterus leucas* (XP_022453898.1); narrow-ridged finless porpoise: *Neophocaena asiaeorientalis* (XP_024596120.1); sperm whale: *Physeter catodon* (XP_007119423.2); Yangtze River dolphin: *Lipotes vexillifer* (XP_007447694.1).

e) northern fur seal: *Callorhinus ursinus* (XP_025723121.1); California sea lion: *Zalophus californianus* (XP_027460417.1); Pacific walrus: *Odobenus rosmarus divergens* (XP_004409206.2);

Weddell seal: *Leptonychotes weddellii* (XP_006747601.1); monk seal: *Neomonachus schauinslandi* (XP_021556282.1).

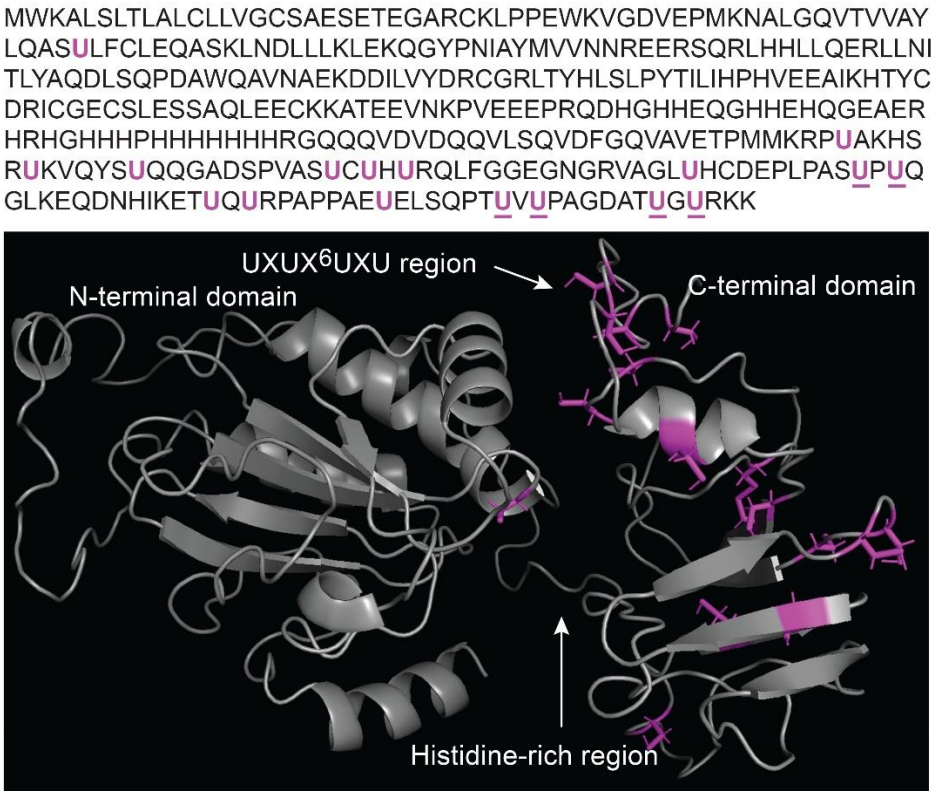


Figure S14. Predicted tertiary structure from I-Tasser[43, 44] of SelP for zebrafish. SelP is made up of a N-terminal domain containing 1 Sec residue (U) and a C-terminal domain containing 16 U residues in zebrafish (and 13 U residues in the grebe). The two domains are connected through a histidine-rich region. The N-terminal domain has redox enzymatic properties and the C-terminal domain supplies Se to extrahepatic tissues[45]. The 17 side chains from the Sec residues in the two domains are represented by purple squares on the left side of the figure. The figure on the right side presents the C-terminal domain. Only the 6 Sec side chains located on the flexible loop after the α-helix are represented. The position of these residues in the amino acid sequence, represented in single-letter code, is denoted with a U symbol. The structures were visualized with PyMol[46].

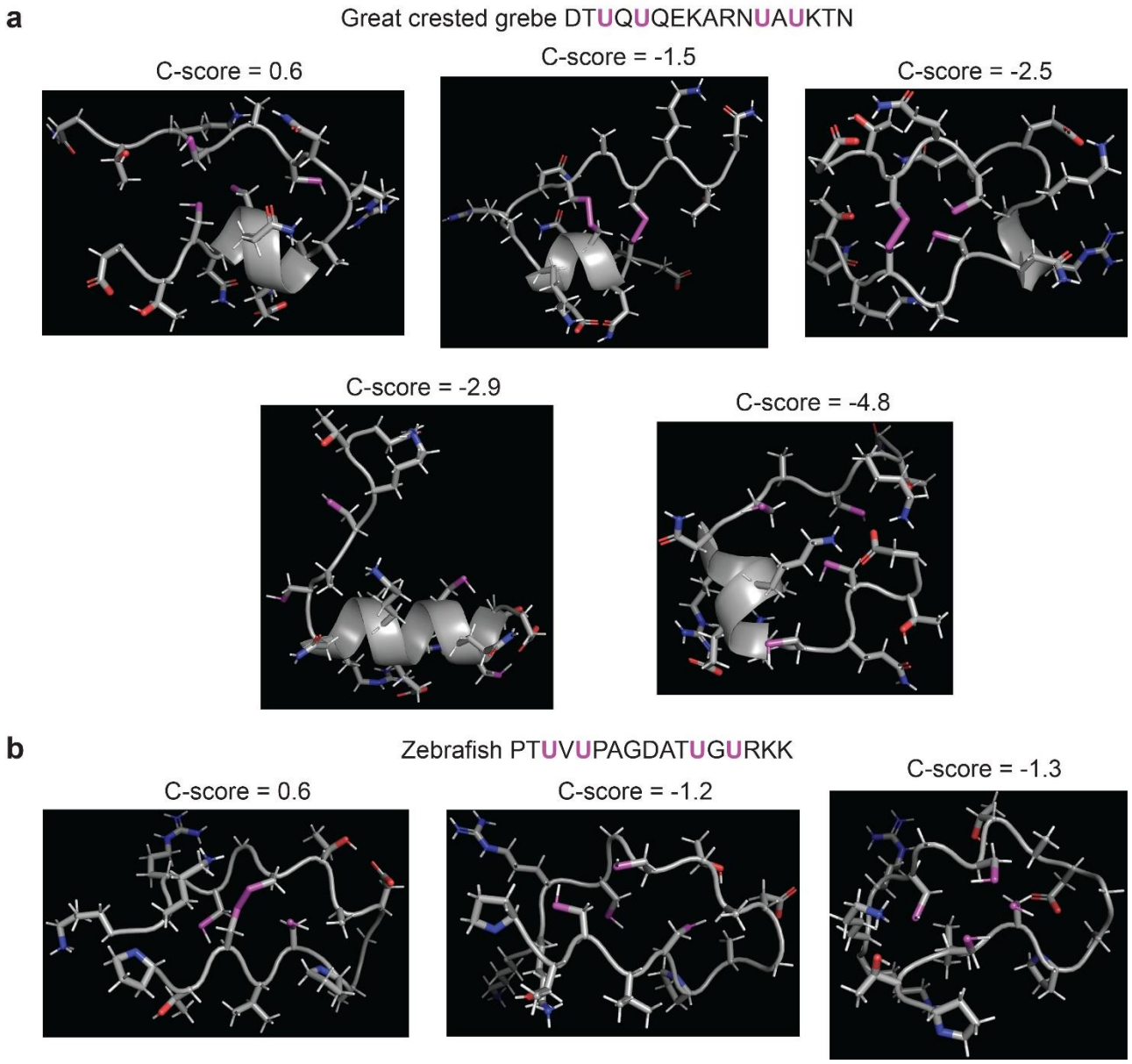


Figure S15. Tertiary structure of the UXUX^{6}UXU region of the C-terminal domain of SelP for the great crested grebe and zebrafish generated by I-Tasser[43, 44]. I-Tasser ranks predicted models according to the C-score confidence parameter and the cluster size. a) Top five predicted models for the great crested grebe. The Sec residues of the three highest C-score models form a cage conducive to binding Hg. Some of the Sec residues are linked through Sec-Sec bridges. b) I-Tasser listed only three models of zebrafish because the top homologous structure templates were similar to each other. This means that the simulation is well converged, thus the structure prediction is of high quality. All conformations were modeled on cysteine-substituted sequences. The amino-acids in each sequence are represented with a single-letter code, with U corresponding to selenocysteine. Purple, red, and blue represent selenol, carboxyl/hydroxyl, and amino groups from amino-acid side chains, respectively.

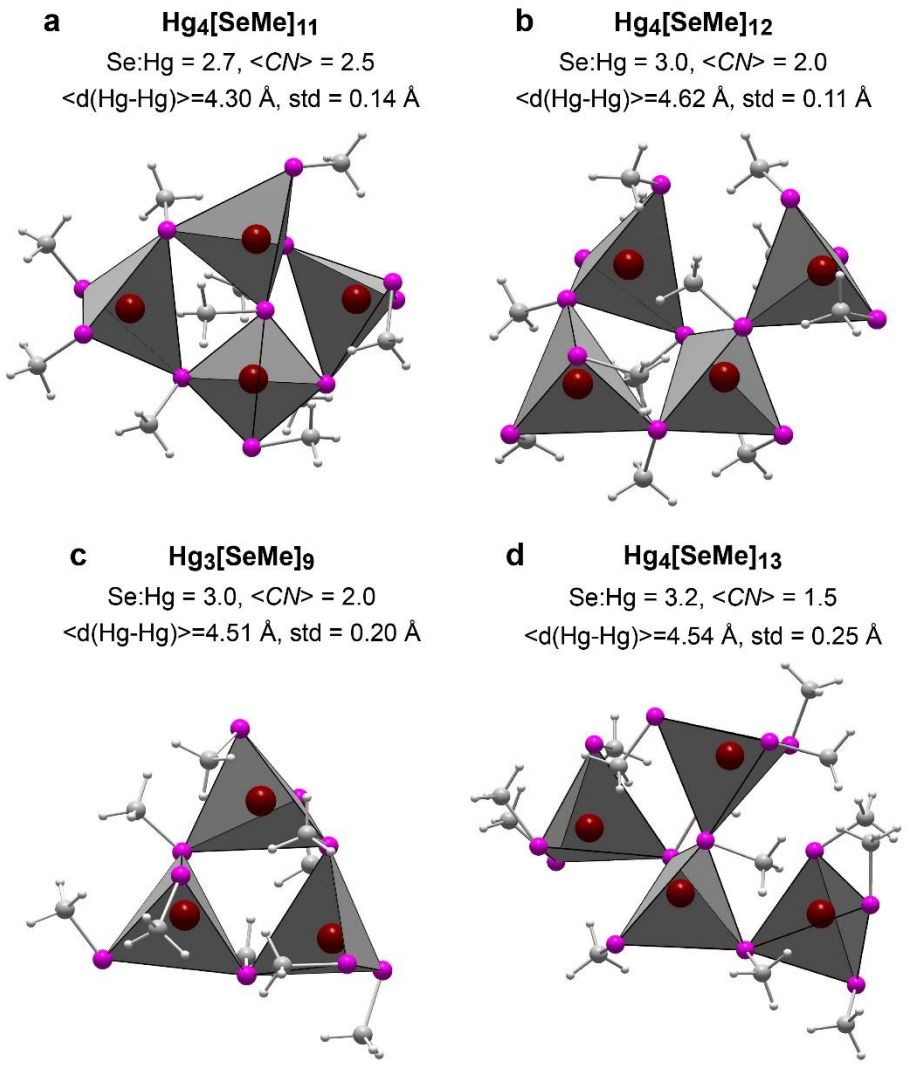


Figure S16. Theoretical topologies of $Hg_x(Se,Sec)_y$ clusters inspired from metallothionein metal clusters with average Hg-Hg coordination numbers (<*CN*>) and distances (<d(Hg-Hg)>). a) $Hg_4(Sec)_{11}$ cluster inspired from the α-$(Zn,Cd)_4(Cys)_{11}$ cluster of metallothionein in vertebrate[47]. b) $Hg_4(Sec)_{12}$ cluster inspired from the $Cd_4(Cys)_{12}$-α cluster of the metallothionein for *Mytilus galloprovincialis*[48]. c,d) $Hg_3(Sec)_9$ and $Hg_4(Sec)_{13}$ clusters inspired from the $(Zn,Cd)_x(Cys)_y$ clusters in bacterial metallothionein[49]. The Hg atom adopts tetrahedral tetraselenolate coordination geometry in all clusters, but the connectivity varies. The Se:Hg ratio can be diminished by substituting a Cys for a Sec ligand. Note that the C-terminal domain of SelP has one Cys residue (Figure S14). Because Hg has an affinity for Se that is approximately two order of magnitude higher than for sulfur[50], the thiol ligand of MeHgCys complexed at the surface of a $Hg_x(Se,Sec)_y$ cluster can be replaced with a selenolate ligand, catalyzing the demethylation of MeHg and promoting the growth of the clusters. All clusters were optimized geometrically at the PBE0-D3 DFT quantum level with def2-TZVP (Hg, Se) and def2-SVP (C, H) basis sets using ORCA 4.2.1[31]. The water solvent was represented with the SMD model[38]. Sec residues were modeled with methaneselenolate (CH_3Se^-). Dark red, Hg; purple, Se; gray, C, light gray, H.

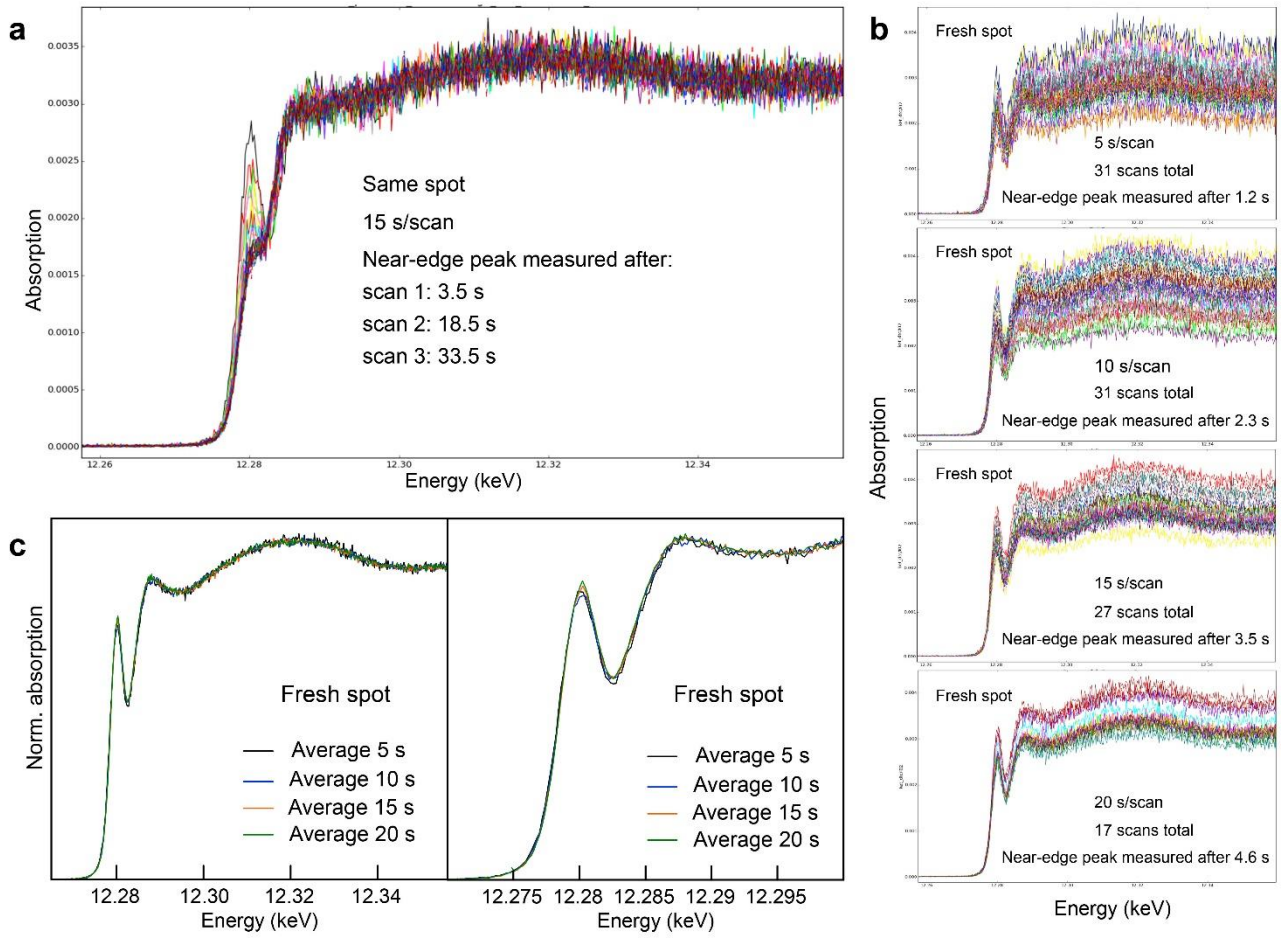


Figure S17. Tests of beam damage. a) The near-edge peak of MeHgCys, which corresponds to transitions of the Hg $2p_{3/2}$ photoelectron to hybridized Hg 6s5d – S 3sp bound-states from the Hg-S bond and to hybridized Hg 6s5d – C 2sp bound-states from the Hg-C bond[32], decreases in intensity after 18.5 s of exposure, and even more so at longer exposure time. The other regions of the HR-XANES spectrum appear stable. In this test, the near-edge peak was first measured (1st scan) after 3.5 s of sample illumination, a duration perhaps already too long to prevent any radiation damage. This possibility was tested next by measuring multiple scans of 5 s, 10 s, 15 s, and 20, each on a fresh spot (b). In this test, the near-edge peak was always measured after 1.2 s (5s scan), 2.3 s (10 s scan), 3.5 s (15 s scan), and 4.6 s (20 s scan) of exposure time. Then each scan series was summed up and normalized to unit step, and the four spectra compared to one another (**c**). Careful evaluation of spectra in **c** shows that the near-edge peak is statistically invariant. We conclude that the samples did not suffer any detectable radiation damage over the 15 s dwell time of all measurements. Furthermore, radiation damage decreases the amplitude of the HR-XANES peaks without changing their energy position, as shown previously in Figure S4 from Ref.[29]. Plots (a) and (b) are screenshots from the PyMCA software[51].

S3. Supplementary tables

Table S1. Body weight, size, and length of fish

Individual	Genus and species	Weight (g)	Body size (cm)[a]
PeacockBass-1	*Cichla monoculus*	356	24.3
PeacockBass-2	*Cichla temensis*	284	24.0
PeacockBass-3	*Cichla temensis*	304	25.5
Pescada-1	*Plagioscion squamosissimus*	225	28.5
Pescada-2	*Plagioscion squamosissimus*	190	27.8

[a]Excluding the caudal fin.

Table S2. Total concentrations of Hg and Se determined by chemical analysis (mg kg^{-1}) and proportions of organic (MeHgCys) and inorganic ($Hg(Sec)_4$ and $Hg(SR)_2$) Hg determined by HR-XANES in fish

Individual and tissue[a]	Genus and species	$[Hg]_{Tot}$ mg kg^{-1}	$[Se]_{Tot}$ mg kg^{-1}	%MeHgCys	$\%Hg(Sec)_4$	$\%Hg(SR)_2$	%Precision	*NSS* x 10^{-5}
PeacockBass-1-M	*Cichla monoculus*	2.8 ± 0.3	0.82 ± 0.08	100	0	0	5	-
PeacockBass-1-L	*Cichla monoculus*	1.2 ± 0.1	3.79 ± 0.38	86	14	0	10	14.8
PeacockBass-2-M	*Cichla temensis*	1.1 ± 0.1	1.06 ± 0.11	100	0	0	5	-
PeacockBass-2-L	*Cichla temensis*	1.7 ± 0.2	7.13 ± 0.71	46	39	16	10	6.1
PeacockBass-3-M	*Cichla temensis*	3.1 ± 0.3	1.01 ± 0.10	100	0	0	5	-
PeacockBass-3-L	*Cichla temensis*	5.5 ± 0.5	9.44 ± 0.94	38	62	0	5	3.2
Pescada-1-M	*Plagioscion squamosissimus*	3.2 ± 0.3	0.81 ± 0.08	100	0	0	5	-
Pescada-1-L	*Plagioscion squamosissimus*	1.4 ± 0.1	1.59 ± 0.16	91	9	0	10	19.5
Pescada-2-M	*Plagioscion squamosissimus*	4.0 ± 0.4	1.10 ± 0.11	100	0	0	5	-
Pescada-2-L	*Plagioscion squamosissimus*	2.0 ± 0.2	2.75 ± 0.27	85	15	0	7	9.5

[a] M denotes muscle tissue and L denotes liver tissue.

Table S3. Total concentrations of Hg and Se determined by chemical analysis (mg kg^{-1}) and proportions of organic (MeHgCys) and inorganic ($Hg(Sec)_4$ and $Hg(SR)_2$) Hg determined by HR-XANES in earthworms

Individual	Sampling site	$[Hg]_{Tot}$ mg kg^{-1}	$[Se]_{Tot}$ mg kg^{-1}	%MeHgSR	$\%Hg(Sec)_4$	$\%Hg(SR)_2$	%Precision	*NSS* x 10^{-5}
EarthW-endo-1 [a]	S1	1.5 ± 0.2	7.51 ± 0.75	25	39	36	10	17.6
EarthW-endo-2	S1	2.4 ± 0.2	7.93 ± 0.79	33	32	34	10	20.3
EarthW-endo-3	S1	1.8 ± 0.2	8.58 ± 0.86	17	10	73	10	10.2
EarthW-endo-4	S1	2.0 ± 0.2	6.76 ± 0.68	30	50	19	10	18.7
EarthW-endo-5	S1	3.1 ± 0.3	4.77 ± 0.48	25	45	28	10	12.2
EarthW-endo-6	S2	4.3 ± 0.4	11.7 ± 1.2	46	39	15	10	10.7
EarthW-endo-7	S2	4.2 ± 0.4	5.88 ± 0.59	30	43	27	10	7.3
EarthW-endo-8	S2	5.1 ± 0.5	28.3 ± 2.8	39	25	36	10	14.7
EarthW-endo-9 [a]	S2	5.4 ± 0.5	26.0 ± 2.6	49	51	0	10	6.0
EarthW-endo-10	S2	10.1 ± 1.0	11.8 ± 1.2	50	50	0	10	9.1
EarthW-anecic-1[a]	S2	1.2 ± 0.1	1.00 ± 0.10	60	29	10	10	12.4

EarthW-anecic-2	S1	1.8 ± 0.2	12.7 ± 1.3	28	72	0	10	5.6
EarthW-anecic-3	S2	5.6 ± 0.6	5.19 ± 0.52	42	44	13	10	7.3

[a]Digestive tract removed.

Table S4. Structural parameters of the Clark's grebe liver derived from Fourier-filtered Hg L_3-edge HR-EXAFS analysis

	Hg-Se			Hg-S			Hg-Hg				
	R (Å)	*CN*	σ^2 (Å)	*R* (Å)	*CN*	σ^2 (Å)	*R* (Å)	*CN*	σ^2 (Å^2)	ΔE_0 (eV)	*Res*
Grebe liver	2.61	2.6	0.008[a]	2.36	0.8	0.008[a]	4.46	2.1	0.01[b]	4.6	5.6[c]
Confidence[d]	0.008	0.6	0.0001	0.026	0.2	0.0001	0.029	0.5	-	1.6	-

[a]Parameters constrained to decrease correlations between σ(Se,S) and *CN*(Se,S). [b]Value fixed to suppress the correlation between σ(Hg) and *CN*(Hg). [c]*Res* = [Σ(| $\chi_{exp} - \chi_{fit}$|)/Σ(| χ_{exp}|)] x 100. [d]Confidence limits of fitting parameters determined from the variation of the fit residual (*Res*) within 95 %.

Table S5. Total concentrations of Hg and Se determined by chemical analysis (mg kg^{-1}) and proportions of organic (MeHgCys) and inorganic ($Hg(Sec)_4$ and $Hg(SR)_2$) Hg determined by HR-XANES in Clark's grebe

Tissue	$[Hg]_{Tot}$ mg kg^{-1}	$[Se]_{Tot}$ mg kg^{-1}	%MeHgCys	%$Hg(Sec)_4$	%$Hg(SR)_2$	%Precision	*NSS* x 10^{-5}
Head feather	-	-	100	0	0	5	-
Breast feather	40.7 ± 4.1	1.04 ± 0.10	100	0	0	5	-
Brain	3.1 ± 0.3	1.55 ± 0.16	100	0	0	5	-
Muscle	6.5 ± 0.7	2.31 ± 0.23	66	11	23	5	1.1
Kidneys	21.6 ± 2.2	10.6 ± 1.0	28	59	12	5	1.9
Liver	43.1 ± 4.3	19.3 ± 1.93	14	86	0	5	2.4

Table S6. Sequences of the oligonucleotides used for PCR amplification

Primers	5'→3' sequences	References
hgcA-F	GGNRTYAAYRTNTGGTGYGC	Ref.[7]
hgcA-912R	GGTGTAGGGGGTGCAGCCSGTRWARKT	Ref.[6]

Table S7. Concentrations of MeHg as determined by HR-XANES and chemical analysis (mg Hg kg^{-1} dry weight)

Individual	[MeHgCys] XANES[a]	Precision	[MeHg] chemical analysis	Precision
Clark's grebe				
Head feather	-	-	-	-
Breast feather	40.7	4.6	34.2	3.4
Brain	3.1	0.3	2.9	0.3
Muscle	4.3	0.5	3.7	0.4
Kidneys	6.0	1.2	6.4	0.6
Liver	6.0	2.2	7.9	0.8
Fish				
PeacockBass-1-M	2.8	0.3	2.0	0.2
PeacockBass-1-L	1.0	0.2	0.85	0.08
PeacockBass-2-M	1.1	0.1	0.81	0.08
PeacockBass-2-L	0.8	0.19	0.45	0.04
PeacockBass-3-M	3.1	0.3	2.2	0.2
PeacockBass-3-L	2.1	0.3	1.8	0.2
Pescada-1-M	3.2	0.3	2.4	0.2
Pescada-1-L	1.3	0.2	0.87	0.09
Pescada-2-M	4.0	0.4	3.0	0.3
Pescada-2-L	1.7	0.2	1.5	0.1
Earthworm				
EarthW-endo-1	0.38	0.16	0.10	0.01
EarthW-endo-2	0.79	0.25	0.32	0.03
EarthW-endo-3	0.31	0.18	0.05	0.01
EarthW-endo-4	0.60	0.21	0.07	0.01
EarthW-endo-5	0.78	0.32	0.14	0.01
EarthW-endo-6	2.0	0.5	1.0	0.1
EarthW-endo-7	1.3	0.4	0.80	0.08
EarthW-endo-8	2.0	0.6	1.3	0.1
EarthW-endo-9	2.6	0.5	1.9	0.2
EarthW-endo-10	5.1	1.1	-	-
EarthW-anecic-1	0.72	0.14	0.39	0.04
EarthW-anecic-2	0.50	0.19	0.14	0.01
EarthW-anecic-3	2.3	0.6	1.3	0.1

[a] $[MeHgCys]_{XANES}$ = %MeHgCys x $[Hg]_{Tot}$.

S4. Supplementary references

(1) Ackerman, J. T.; Eagles-Smith, C. A.; Herzog, M. P.; Hartman, C. A.; Peterson, S. H.; Evers, D. C.; Jackson, A. K.; Elliott, J. E.; Vander Pol, S. S.; Bryan, C. E. Avian mercury exposure and toxicological risk across western North America: A synthesis. *Sci. Tot. Environ.* **2016,** *568*, 749-769.

(2) Kleckner, A. E.; Kakouros, E.; Stewart, A. R. A practical method for the determination of total selenium in environmental samples using isotope dilution-hydride generation-inductively coupled plasma-mass spectrometry. *Limnol. Oceanogr.-Meth.* **2017,** *15*, 363-371.

(3) Gajdosechova, Z.; Mester, Z.; Feldmann, J.; Krupp, E. M. The role of selenium in mercury toxicity - Current analytical techniques and future trends in analysis of selenium and mercury interactions in biological matrices. *Trac-Trend. Anal. Chem.* **2018,** *104*, 95-109.

(4) Byun, M. O. K.; Kaper, J. B.; Ingram, L. O. Construction of a new vector for the expression of foreign genes in *Zymomonas mobilis*. *J. Ind. Microbiol.* **1986,** *1*, 9-15.

(5) Parks, J. M.; Johs, A.; Podar, M.; Bridou, R.; Hurt, R. A.; Smith, S. D.; Tomanicek, S. J.; Qian, Y.; Brown, S. D.; Brandt, C. C.; Palumbo, A. V.; Smith, J. C.; Wall, J. D.; Elias, D. A.; Liang, L. Y. The genetic basis for bacterial mercury methylation. *Science* **2013,** *339*, 1332-1335.

(6) Schaefer, J. K.; Kronberg, R. M.; Morel, F. M. M.; Skyllberg, U. Detection of a key Hg methylation gene, *hgcA*, in wetland soils. *Env. Microbiol. Rep.* **2014**, 441-447.

(7) Bae, H. S.; Dierberg, F. E.; Ogram, A. Syntrophs dominate sequences associated with the mercury methylation-related gene *hgcA* in the water conservation areas of the Florida Everglades. *Appl. Environ. Microb.* **2014,** *80*, 6517-6526.

(8) Liu, Y. R.; Yu, R. Q.; Zheng, Y. M.; He, J. Z. Analysis of the microbial community structure by monitoring an Hg methylation gene (*hgcA*) in paddy soils along an Hg gradient. *Appl. Environ. Microb.* **2014,** *80*, 2874-2879.

(9) George, G. N.; Pickering, I. J.; Pushie, M. J.; Nienaber, K.; Hackett, M. J.; Ascone, I.; Hedman, B.; Hodgson, K. O.; Aitken, J. B.; Levina, A.; Glover, C.; Lay, P. A. X-ray-induced photo-chemistry and X-ray absorption spectroscopy of biological samples. *J. Synchrotron Radiat.* **2012,** *19*, 875-886.

(10) Manceau, A.; Bustamante, P.; Haouz, A.; Bourdineaud, J. P.; Gonzalez-Rey, M.; Geertsen, V.; Barruet, E.; Rovezzi, M.; Glatzel, P.; Pin, S. Mercury(II) binding to metallothionein in *Mytilus edulis* revealed by high energy-resolution XANES spectroscopy. *Chem-Eur J.* **2019,** *25*, 997-1009.

(11) Rovezzi, M.; Lapras, C.; Manceau, A.; Glatzel, P.; Verbeni, R. High energy-resolution x-ray spectroscopy at ultra-high dilution with spherically bent crystal analyzers of 0.5 m radius. *Rev. Sci. Instr.* **2017,** *88*, 013108.

(12) Manceau, A.; Marcus, M. A.; Tamura, N. Quantitative speciation of heavy metals in soils and sediments by synchrotron X-ray techniques. In *Applications of Synchrotron Radiation in Low-Temperature Geochemistry and Environmental Science*, Fenter, P. A.; Rivers, M. L.; Sturchio, N. C.; Sutton, S. R., Eds. Mineralogical Society of America: Washington, DC, 2002; Vol. 49, pp 341-428.

(13) Manceau, A.; Marcus, M.; Lenoir, T. Estimating the number of pure chemical components in a mixture by X-ray absorption spectroscopy. *J. Synchrotron Radiat.* **2014,** *21*, 1140-1147.

(14) Rossberg, A.; Reich, T.; Bernhard, G. Complexation of uranium(VI) with protocatechuic acid-application of iterative transformation factor analysis to EXAFS spectroscopy. *Anal. Bioanal. Chem.* **2003,** *376*, 631-638.

(15) Malinowski, E. R. Theory of error for target factor-analysis with applications to mass-spectrometry and nuclear magnetic-resonance spectrometry. *Anal. Chim. Acta-Comp.* **1978,** *2*, 339-354.

(16) Mah, V.; Jalilehvand, F. Glutathione complex formation with mercury(II) in aqueous solution at physiological pH. *Chem. Res. Toxicol.* **2010,** *23*, 1815-1823.

(17) Manceau, A.; Wang, J.; Rovezzi, M.; Glatzel, P.; Feng, X. Biogenesis of mercury-sulfur nanoparticles in plant leaves from atmospheric gaseous mercury. *Environ. Sci. Technol.* **2018,** *52*, 3935–3948.

(18) Bourdineaud, J. P.; Gonzalez-Rey, M.; Rovezzi, M.; Gatzel, P.; Nagy, K. L.; Manceau, A. Divalent mercury from dissolved organic matter is bioavailable to fish and accumulates as dithiolate and tetrathiolate complexes. *Environ. Sci. Technol.* **2019,** *53*, 4880−4891.

(19) Marcus, M. A.; MacDowell, A. A.; Celestre, R.; Manceau, A.; Miller, T.; Padmore, H. A.; Sublett, R. E. Beamline 10.3.2 at ALS: a hard X-ray microprobe for environmental and materials sciences. *J. Synchrotron Radiat.* **2004,** *11*, 239-247.

(20) Ressler, T. WinXAS: a program for X-ray absorption spectroscopy data analysis under MS-Windows. *J. Synchrotron Radiat.* **1998,** *5*, 118-122.

(21) Ankudinov, A. L.; Rehr, J. J. Relativistic calculations of spin-dependent X-ray-absorption spectra. *Phys. Rev. B* **1997,** *56*, 1712-1716.

(22) Earley, J. W. Description and synthesis of the selenide minerals. *Am. Miner.* **1950,** *35*, 337-364.

(23) Roman, M.; Sanchez, M. L. F.; Sanz-Medel, A.; Iglesias, H. G.; Cescon, P.; Barbante, C. Selenium speciation in rat colon tissues. *J. Anal. Atom. Spectrom.* **2011,** *26*, 100-108.

(24) Busto, M. E. D.; Oster, C.; Cuello-Nunez, S.; Deitrich, C. L.; Raab, A.; Konopka, A.; Lehmann, W. D.; Goenaga-Infante, H.; Fisicaro, P. Accurate quantification of selenoproteins in human plasma/serum by isotope dilution ICP-MS: focus on selenoprotein P. *J. Anal. At. Spectrom.* **2016,** *31*, 1904-1912.

(25) Sandalova, T.; Zhong, L. W.; Lindqvist, Y.; Holmgren, A.; Schneider, G. Three-dimensional structure of a mammalian thioredoxin reductase: Implications for mechanism and evolution of a selenocysteine-dependent enzyme. *Proc. Natl. Acad. Sci. U.S.A.* **2001,** *98*, 9533-9538.

(26) Pickering, I. J.; Cheng, Q.; Rengifo, E. M.; Nehzati, S.; Dolgova, N. V.; Kroll, T.; Sokaras, D.; George, G. N.; Arner, E. S. J. Direct observation of methylmercury and auranofin binding to selenocysteine in thioredoxin reductase. *Inorg Chem.* **2020,** *59*, 2711-2718.

(27) Ren, B.; Huang, W. H.; Akesson, B.; Ladenstein, R. The crystal structure of seleno-glutathione peroxidase from human plasma at 2.9 Å resolution. *J. Mol. Biol.* **1997,** *268*, 869-885.

(28) Manceau, A.; Lemouchi, C.; Enescu, M.; Gaillot, A.-C.; Lanson, M.; Magnin, V.; Glatzel, P.; Poulin, B. A.; Ryan, J. N.; Aiken, G. R.; Gautier-Luneau, I.; Nagy, K. L. Formation of mercury sulfide from Hg(II)-thiolate complexes in natural organic matter. *Environ. Sci. Technol.* **2015,** *49*, 9787−9796.

(29) Manceau, A.; Enescu, M.; Simionovici, A.; Lanson, M.; Gonzalez-Rey, M.; Rovezzi, M.; Tucoulou, R.; Glatzel, P.; Nagy, K. L.; Bourdineaud, J.-P. Chemical forms of mercury in human hair reveal sources of exposure. *Environ. Sci. Technol.* **2016,** *50*, 10721–10729.

(30) Møller, C.; Plesset, M. S. Note on an approximation treatment for many-electron systems. *Phys. Rev.* **1934,** *46*, 618-622.

(31) Neese, F. The ORCA program system. *WIREs Comput. Mol. Sci.* **2012,** *2*, 73-78.

(32) Manceau, A.; Lemouchi, C.; Rovezzi, M.; Lanson, M.; Glatzel, P.; Nagy, K. L.; Gautier-Luneau, I.; Joly, Y.; Enescu, M. Structure, bonding, and stability of mercury complexes with thiolate and thioether ligands from high-resolution XANES spectroscopy and first-principles calculations. *Inorg. Chem.* **2015,** *54*, 11776−11791.

(33) Schafer, A.; Horn, H.; Ahlrichs, R. Fully optimized contracted Gaussian-basis sets for atoms Li to Kr. *J. Chem. Phys.* **1992,** *97*, 2571-2577.

(34) Weigend, F.; Ahlrichs, R. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Phys. Chem. Chem. Phys.* **2005,** *7*, 3297-3305.

(35) Feyereisen, M.; Fitzgerald, G.; Komornicki, A. Use of approximate integrals in ab initio theory - An application in MP2 energy calculations. *Chem. Phys. Lett.* **1993,** *208*, 359-363.

(36) Kendall, R. A.; Fruchtl, H. A. The impact of the resolution of the identity approximate integral method on modern ab initio algorithm development. *Theor. Chem. Acc.* **1997,** *97*, 158-163.

(37) Haussermann, U.; Dolg, M.; Stoll, H.; Preuss, H.; Schwerdtfeger, P.; Pitzer, R. M. Accuracy of energy-adjusted quasi-relativistic *ab initio* pseudopotentials - All electron and pseudopotential benchmark calculations for Hg, HgH and their cations. *Mol. Phys.* **1993,** *78*, 1211-1224.

(38) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions. *Phys. Chem. B* **2009,** *113*, 6378-6396.

(39) Cattell, R. B. The scree test for the number of factors. *J. Multiv. Behav. Res.* **1996,** *1*, 245–276.
(40) Rehr, J. J.; Mustre de Leon, J.; Zabinsky, S. I.; Albers, R. C. Theoretical X-ray absorption fine structure standards. *J. Am. Chem. Soc.* **1991,** *113*, 5135-5145.
(41) Teo, B. K. *EXAFS: basic principles and data analysis*. Springer-Verlag: Berlin, 1986; p 349.
(42) Crozier, E. D. A review of the current status of XAFS spectroscopy. *Nucl. Instrum. Meth. B* **1997,** *133*, 134-144.
(43) Roy, A.; Kucukural, A.; Zhang, Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nature Protoc.* **2010,** *5*, 725-738.
(44) Yang, J.; Yan, R.; Roy, A.; Xu, D.; Poisson, J.; Zhang, Y. The I-TASSER Suite: Protein structure and function prediction. *Nat. Methods* **2015,** *12*, 7-8.
(45) Burk, R. F.; Hill, K. E. Regulation of selenium metabolism and transport. In *Annu. Rev. Nutr.*, Bowman, B. A.; Stover, P. J., Eds. 2015; Vol. 35, pp 109-134.
(46) DeLano, W. L. The PyMOL Molecular Graphics System. *http://www.pymol.org* **2002**.
(47) Hidalgo, J.; Chung, R. S.; Penkowa, M.; Vasak, M. Structure and function of vertebrate metallothioneins. *Met. Ions Life Sci.* **2009,** *5*, 279–317.
(48) Digilio, G.; Bracco, C.; Vergani, L.; Botta, M.; Osella, D.; Viarengo, A. The cadmium binding domains in the metallothionein isoform Cd-7-MT10 from *Mytilus galloprovincialis* revealed by NMR spectroscopy. *J. Biol. Inorg. Chem.* **2009,** *14*, 167-178.
(49) Habjanic, J.; Zerbe, O.; Freisinger, E. A histidine-rich *Pseudomonas* metallothionein with a disordered tail displays higher binding capacity for cadmium than zinc. *Metallomics* **2018,** *10*, 1415-1429.
(50) Ralston, N. V. C.; Raymond, L. J. Mercury's neurotoxicity is characterized by its disruption of selenium biochemistry. *BBA-Gen. Sub.* **2018,** *1862*, 2405-2416.
(51) Sole, V. A.; Papillon, E.; Cotte, M.; Walter, P.; Susini, J. A multiplatform code for the analysis of energy-dispersive X-ray fluorescence spectra. *Spectrochim. Acta, B* **2007,** *62*, 63-68.

S5. Cartesian coordinates of $Hg(Selenoneine)_4$ and $Hg_{10}(SeMe)_{20}$

Hg[Selenoneine]4				Hg10(SeMe)20			
Hg	3.461701	4.146748	5.583562	Hg	6.014420	-0.058675	0.371545
Se	2.706204	2.013233	6.875411	Hg	-0.268486	6.515726	0.411108
Se	4.922708	4.001610	3.333475	Hg	-0.253160	0.430061	6.310613
Se	5.232110	5.278112	7.089261	Hg	5.938762	6.269286	6.175340
Se	1.193700	5.029105	4.657621	Hg	-0.155567	3.163920	3.116928
C	4.476669	1.607680	7.394439	Hg	5.984914	3.240218	3.043747
N	5.423897	1.224327	6.487590	Hg	2.548632	0.020246	2.977366
N	4.988642	1.697864	8.622831	Hg	3.151103	6.329468	2.859888
C	6.614491	1.079474	7.153129	Hg	2.920073	3.173796	-0.766157
C	6.314848	1.369774	8.480903	Hg	3.197494	3.087935	6.342012
C	7.870810	0.569132	6.523838	Se	1.444574	1.863416	1.431035
H	7.004447	1.354284	9.326655	Se	1.795140	4.497964	4.416517
H	5.237060	1.111153	5.464794	Se	4.350104	1.597700	4.401854
C	8.245636	1.102636	5.132285	Se	4.441839	4.573375	1.325447
H	7.771604	-0.513782	6.363167	Se	7.554977	1.392694	1.932403
H	8.701825	0.700246	7.234270	Se	7.328923	-1.522043	-1.319519
C	9.100078	0.031958	4.365074	Se	4.936997	1.842911	-1.452448
O	8.599791	-1.113986	4.414527	Se	3.892732	-1.393875	1.209287
O	10.145865	0.379327	3.760442	Se	1.814804	7.912470	1.251024
N	8.839878	2.507015	5.144193	Se	0.771487	4.391640	-1.227576
C	10.186464	2.544438	5.788929	Se	-1.579017	7.688114	-1.476559
C	7.936086	3.437658	5.889650	Se	-1.640263	5.149015	2.193377
C	8.935495	3.027086	3.741204	Se	1.106600	2.250149	7.660173
H	10.089642	2.212864	6.827122	Se	1.296942	-1.302379	4.801895
H	10.855192	1.891119	5.225267	Se	-1.777931	1.586978	4.434928
H	10.534462	3.582794	5.765039	Se	-1.487740	-1.257628	7.833693
H	6.928722	3.365774	5.463529	Se	7.688998	7.743364	7.381857

H	7.926431	3.167904	6.948110	Se	7.506932	4.885899	4.372863
H	8.327102	4.453916	5.761155	Se	4.016526	7.617318	4.974804
H	9.616915	2.388822	3.179986	Se	5.038407	4.267409	7.739551
H	7.930301	3.000866	3.307192	C	-2.577701	6.166898	-2.190697
H	9.299819	4.059148	3.791551	C	-2.763118	4.081739	1.008201
H	7.325735	1.209258	4.536808	C	2.866505	7.640337	-0.365368
C	5.314390	2.207598	2.909719	C	5.015315	9.102027	4.205908
N	6.062667	1.888638	1.816692	C	5.702831	5.550858	0.199722
N	4.938404	1.086620	3.544192	C	4.078636	0.522030	-2.601745
C	6.192786	0.524961	1.740697	C	0.316836	0.770403	0.277591
C	5.473034	0.041798	2.824503	C	8.774283	6.280457	8.094170
C	6.933547	-0.178036	0.650556	C	0.718600	5.677103	5.530074
H	5.334394	-0.996833	3.123961	C	7.533846	6.371537	3.113138
H	6.606628	2.507978	1.205970	C	6.502881	3.031870	7.392416
C	8.461269	0.049332	0.686000	C	-1.827617	0.079819	3.201773
H	6.534300	0.108159	-0.335675	C	5.670190	0.352466	5.112999
H	6.697426	-1.240994	0.778198	C	1.939028	1.144315	9.032010
C	8.853129	1.276894	-0.196569	C	-2.438132	-2.211425	6.415179
O	8.286205	2.339358	0.164199	C	2.724464	-1.425207	6.120654
O	9.647899	1.144029	-1.151784	C	4.724793	-2.714985	2.377020
N	9.229358	-1.236191	0.387777	C	8.407714	2.533342	0.598741
C	8.997964	-1.738080	-0.998742	C	9.076558	-0.652161	-1.210952
C	8.825028	-2.307698	1.355052	C	1.197052	5.351923	-2.868638
C	10.689050	-0.985822	0.618930	H	-1.905783	5.331305	-2.426378
H	7.922989	-1.895179	-1.137591	H	-3.068344	6.499632	-3.115035
H	9.376887	-0.996577	-1.702159	H	-3.343099	5.832244	-1.480246
H	9.527322	-2.692370	-1.099164	H	2.160965	5.865415	-2.782145
H	8.785675	-1.877118	2.364424	H	1.215043	4.648069	-3.708828
H	7.853047	-2.716692	1.068446	H	0.402830	6.097368	-3.013214
H	9.576852	-3.101977	1.295872	H	3.320857	-0.042712	-2.046112
H	11.018791	-0.203152	-0.064298	H	3.625598	1.023837	-3.464629
H	10.801274	-0.677730	1.664451	H	4.870700	-0.159148	-2.939445
H	11.224123	-1.922454	0.429380	H	9.412828	-0.587826	-0.168218
H	8.769237	0.281483	1.715532	H	9.036973	0.349996	-1.655232
C	4.072920	6.182070	8.280140	H	9.791109	-1.264360	-1.777040
N	2.999383	6.906641	7.848804	H	7.700523	2.804248	-0.194013
N	4.169523	6.215033	9.613408	H	9.254966	1.989107	0.165808
C	2.358402	7.421155	8.946524	H	8.778477	3.434425	1.103583
C	3.113676	6.986790	10.033404	H	6.509284	6.668405	2.860436
C	1.030548	8.092146	8.832323	H	8.079593	6.078526	2.208510
H	2.923583	7.180587	11.090468	H	8.054374	7.207914	3.596685
H	2.764584	7.122028	6.848721	H	9.658845	6.718935	8.574949
C	-0.093881	7.252069	8.213801	H	9.101352	5.615642	7.284195
H	1.106642	8.966946	8.163711	H	8.209649	5.708970	8.841960
H	0.715493	8.485069	9.807664	H	5.840712	8.745859	3.580634
C	-1.358582	8.136822	8.015376	H	4.322785	9.705796	3.606506
O	-1.982334	7.958819	6.942424	H	5.408324	9.705104	5.033794
O	-1.630013	8.944077	8.940645	H	2.951063	6.567454	-0.574571
N	-0.463185	5.979873	8.982234	H	3.857814	8.087540	-0.225905
C	-0.602789	6.239615	10.443706	H	2.349959	8.141534	-1.193229
C	0.552443	4.902638	8.743772	H	-3.543708	4.732072	0.595754
C	-1.765326	5.453439	8.462592	H	-2.169169	3.642667	0.197890
H	0.387067	6.445911	10.860475	H	-3.230954	3.287788	1.604503
H	-1.269955	7.096797	10.576328	H	-2.298219	0.400382	2.264552
H	-1.021166	5.338885	10.906317	H	-0.815419	-0.296865	3.012353
H	0.649966	4.774111	7.658422	H	-2.431511	-0.713754	3.658293
H	1.507475	5.188258	9.186681	H	-2.951959	-3.068435	6.870542
H	0.180134	3.985629	9.214665	H	-3.184353	-1.559457	5.943538

H	-2.571333	6.129948	8.759431
H	-1.701657	5.388058	7.372097
H	-1.921706	4.464143	8.905460
H	0.206007	6.878728	7.225946
C	1.384638	6.901459	4.696856
N	0.441676	7.748462	4.205295
N	2.395108	7.598284	5.230934
C	0.839380	9.037902	4.444784
C	2.065257	8.926150	5.085920
C	-0.009938	10.208516	4.073822
H	2.714215	9.727207	5.440882
H	-0.479433	7.549914	3.785547
C	-1.440305	10.056503	4.642710
H	-0.052853	10.331025	2.979355
H	0.490467	11.107074	4.456184
C	-2.370958	9.503238	3.527842
O	-2.051544	8.338502	3.167480
O	-3.292554	10.205034	3.059991
N	-1.934200	11.308615	5.346992
C	-1.962584	12.499479	4.448331
C	-1.055172	11.596550	6.522633
C	-3.303404	11.040224	5.892413
H	-0.937738	12.739649	4.148426
H	-2.578651	12.252913	3.581995
H	-2.388122	13.336519	5.012871
H	-1.731876	-2.579941	5.659929
H	1.137887	0.719148	9.649280
H	2.532136	0.338636	8.586185
H	2.579041	1.785807	9.650208
H	7.404312	3.432219	7.872218
H	6.674932	2.939355	6.313917
H	6.260575	2.054166	7.825434
H	0.049655	6.262629	4.888000
H	1.383194	6.348129	6.086926
H	0.130640	5.068963	6.227798
H	5.219987	5.751736	-0.763293
H	6.587330	4.924801	0.038130
H	5.994666	6.492895	0.673539
H	5.284887	-0.112503	6.026284
H	5.856528	-0.416584	4.354237
H	6.601344	0.886101	5.329208
H	3.926817	-3.285435	2.868528
H	5.330827	-3.388222	1.758390
H	5.360154	-2.234158	3.128720
H	2.321972	-1.949164	6.996786
H	3.558620	-2.001895	5.702565
H	3.059782	-0.422316	6.409620
H	-0.314238	1.437707	-0.321754
H	-0.304458	0.094089	0.871688
H	0.969588	0.185795	-0.382365
H	-1.113486	10.747463	7.212104
H	-0.027473	11.761020	6.189701
H	-1.433117	12.504484	7.005355
H	-3.990983	10.908874	5.056052
H	-3.237230	10.131552	6.501737
H	-3.592522	11.896897	6.510645
H	-1.431745	9.303615	5.444663