Ancient Loss of Catalytic Selenocysteine Spurred Convergent Adaptation in a Mammalian Oxidoreductase

Jasmin Rees [1,2,*,†], Gaurab Sarangi[3,†], Qing Cheng[4,†], Martin Floor[5,6,†], Aida M Andrés[2], Baldomero Oliva Miguel [7], Jordi Villà-Freixa [5,8], Elias S.J. Arnér[4,9], and Sergi Castellano [1,10,*]

[1]Great Ormond Street Institute of Child Health, University College London, London, UK
[2]Division of Biosciences, University College London, London, UK
[3]Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany
[4]Division of Biochemistry, Department of Medical Biochemistry and Biophysics, Karolinska Institutet, Stockholm, Sweden
[5]Department of Biosciences, Faculty of Sciences and Technology, Universitat de Vic—Universitat Central de Catalunya, Vic, Spain
[6]Department of Life Sciences, Barcelona Supercomputing Center (BSC), Barcelona, Spain
[7]Department of Health and Experimental Sciences, Universitat Pompeu Fabra, Barcelona, Spain
[8]Institut de Recerca i Innovació en Ciències de la Vida i de la Salut a la Catalunya Central (IRIS-CC), Vic, Spain
[9]Department of Selenoprotein Research, National Institute of Oncology, Budapest, Hungary
[10]UCL Genomics, University College London, London, UK

[†]These authors contributed equally.

*Corresponding authors: E-mails: jasmin.rees.18@ucl.ac.uk; s.castellano@ucl.ac.uk.

Accepted: February 22, 2024

Abstract

Selenocysteine, the 21st amino acid specified by the genetic code, is a rare selenium-containing residue found in the catalytic site of selenoprotein oxidoreductases. Selenocysteine is analogous to the common cysteine amino acid, but its selenium atom offers physical–chemical properties not provided by the corresponding sulfur atom in cysteine. Catalytic sites with selenocysteine in selenoproteins of vertebrates are under strong purifying selection, but one enzyme, glutathione peroxidase 6 (GPX6), independently exchanged selenocysteine for cysteine <100 million years ago in several mammalian lineages. We reconstructed and assayed these ancient enzymes before and after selenocysteine was lost and up to today and found them to have lost their classic ability to reduce hydroperoxides using glutathione. This loss of function, however, was accompanied by additional amino acid changes in the catalytic domain, with protein sites concertedly changing under positive selection across distant lineages abandoning selenocysteine in glutathione peroxidase 6. This demonstrates a narrow evolutionary range in maintaining fitness when sulfur in cysteine impairs the catalytic activity of this protein, with pleiotropy and epistasis likely driving the observed convergent evolution. We propose that the mutations shared across distinct lineages may trigger enzymatic properties beyond those in classic glutathione peroxidases, rather than simply recovering catalytic rate. These findings are an unusual example of adaptive convergence across mammalian selenoproteins, with the evolutionary signatures possibly representing the evolution of novel oxidoreductase functions.

Key words: selenocysteine, convergent, selenoprotein, catalysis, adaptation.

Significance

Selenocysteine (Sec), the 21st amino acid, offers unique physical–chemical properties but has been independently lost in the GPX6 protein in several mammalian lineages. By reconstructing ancient mammalian proteins of GPX6, we show that exchanging Sec for its analogous amino acid cysteine (Cys) in this protein results in a loss of traditional catalytic function that is not regained over evolutionary time but is also accompanied by amino acid exchanges that are enriched for signatures of positive selection and shared across disparate lineages. Hence, we demonstrate a narrow evolutionary range following the Sec-to-Cys exchange and propose adaptive convergence of this protein across multiple mammalian lineages.

Introduction

Catalytic residues are largely conserved in enzymes (Sharir-Ivry and Xia 2021) as they lower the activation energy of reactions and thereby can increase enzymatic turnover. Mutations in these evolutionarily constrained active sites typically reduce catalytic activity (Carter and Wells 1988; Loeb et al. 1989; Rennell et al. 1991), a proxy for fitness in enzymes, and are consequently often deleterious. Still, when such mutations do occur and persist, proteins often demonstrate evolutionary trajectories which either recover the rate of catalysis (Gromer et al. 2003) or open new protein functions (Jensen 1976; Jayaraman et al. 2022). Mutations that recover fitness, termed compensatory mutations, may not necessarily occur after a deleterious mutation to directly increase fitness but may also precede a deleterious mutation to prevent the loss of fitness (Jayaraman et al. 2022).

The evolutionary trajectories that follow a deleterious mutation, however, are limited by the enzyme's sequence, as natural selection favors mutations whose interactions with other residues compensate for catalytic loss or advance alternative properties, thereby increasing the fitness of the protein. Alongside epistatic interactions dictating the fitness of individual mutations, pleiotropy also plays a key role: trajectories that improve one enzymatic function but compromise another can decrease the fitness of a mutation and may be under purifying selection (Weinreich et al. 2006; Storz 2016). Such epistatic and pleiotropic constraints are often similar in orthologous proteins, whose sequence conservation among species provides similar genetic backgrounds for mutations and may result in natural selection favoring the evolution of the same sites across species.

Here, we study the selenoprotein glutathione peroxidase 6 ($GPX6_{Sec}$), present only in placental mammals. We focus on the sporadic replacement, in several lineages, of the rare amino acid selenocysteine (Sec) for cysteine (Cys) (Kryukov et al. 2003) and its contrast to proteins of the GPX family using exclusively either Sec (GPX1, 2, 3, and 4) or Cys (GPX5, 7, and 8). Sec-to-Cys substitutions across orthologous selenoproteins, as seen in mammalian $GPX6_{Sec}$, are unusual (Castellano et al. 2005) with other GPX_{Cys} proteins emerging only after duplication of a GPX_{Sec} gene (Fig. 1a; Mariotti et al. 2012). This is due to the low exchangeability of Sec and Cys in catalysis, likely due to the deleterious nature of the Sec-to-Cys mutations. Strong purifying selection (Castellano et al. 2009) is believed to act against the lower catalytic activity, lower nucleophilicity, and lower efficiency as a leaving group of Cys compared with Sec (Axley et al. 1991; Berry et al. 1992; Lee et al. 2000; Johansson et al. 2005; Arnér 2010; Kim et al. 2015; Reich and Hondal 2016), rendering these exchanges rare and likely opposed by natural selection.

Classic GPX_{Sec} activity reduces hydroperoxides, particularly hydrogen and lipid peroxides, with glutathione (GSH) as a cofactor. GPX_{Cys} proteins, from early duplications in early history of mammals ($GPX5_{Cys}$ from $GPX3_{Sec}$ duplication; around 300 million years ago [Mya]), vertebrates ($GPX8_{Cys}$ from $GPX7_{Cys}$ or $GPX4_{Sec}$ duplication; probably 450 Mya), and metazoan ($GPX7_{Cys}$ from $GPX4_{Sec}$ duplication; more than 1,000 Mya) (Hedges 2002; Castellano et al. 2009; Trenz et al. 2021), have evolved a preference for other cofactors, for example, thioredoxin in $GPX5_{Cys}$ or protein disulfide isomerase (PDI) in $GPX7_{Cys}$ and $GPX8_{Cys}$ (Nguyen et al. 2011). These Cys-containing proteins act on alternative substrates for peroxidation and may have additional functions, including signaling and oxidative protein folding (Nguyen et al. 2011; Taylor et al. 2013; Buday and Conrad 2021). Thus, while all GPX proteins may protect cells from oxidative damage (Tosatto et al. 2008), those proteins containing Cys may employ novel pathways to do so without being bona fide peroxidases, perhaps on account of their lower catalytic turnover. We ask if this is also the case for $GPX6_{Cys}$.

Results

Increased Rate of Evolution Surrounding the Loss of Sec

We inferred 5 independent losses of Sec in $GPX6_{Sec}$ (Fig. 1b, dashed green branches) across 22 mammals by reconstructing the ancestral sequence at each node of their phylogeny with PAML (supplementary fig. S1, Supplementary Material online; Yang 2007). These inferred losses occurred in the last 64 million years (Fig. 1b, dashed green branches; Huchon et al. 2002; Steppan et al. 2004; Higdon et al. 2007; Hallström and Janke 2008; Chatterjee et al. 2009; Nyakatura and Bininda-Emonds 2012) and resulted in

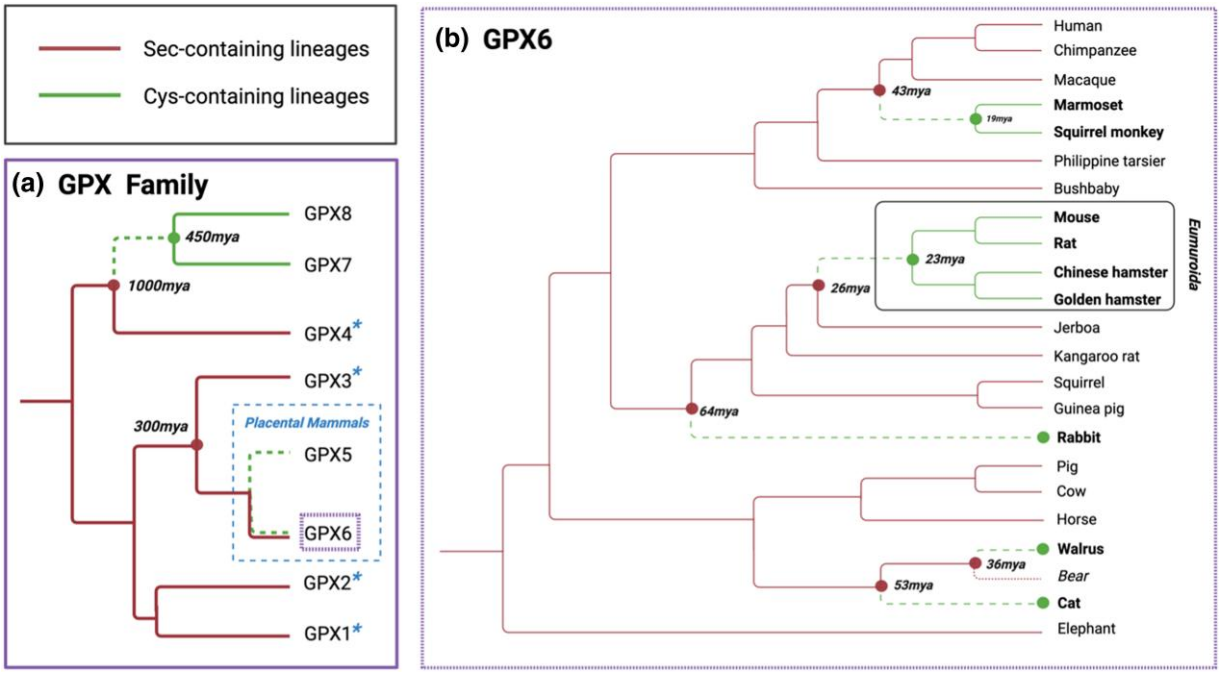

Fig. 1.—Phylogenies of the GPX family. a) The phylogeny of the GPX family in Eukaryotes, with selenoproteins present in the vertebrate ancestor indicated with a blue asterisk (based on Mariotti et al. (2012)). Includes the dates of the duplications leading to $GPX7_{Cys}$, $GPX8_{Cys}$, and $GPX5_{Cys}$ and their older, single substitutions of Sec to Cys that resulted in enzymes with new properties. GPX5 and GPX6 are only present in placental mammals, indicated by the dashed blue box. b) The topology of the phylogeny of the 22 mammals in our analysis. In red, $GPX6_{Sec}$ branches, in green, $GPX6_{Cys}$ ones. Dashed green branches represent $GPX6_{Cys}$ lineages where Sec was lost. Dotted red branch indicates the bear $GPX6_{Sec}$ lineage, which was not used in the analysis due to sequence quality issues. The $GPX6_{Cys}$ *Eumuroida* clade, a specific group of muroid rodents, is boxed. Approximate ages given by Huchon et al. (2002), Steppan et al. (2004), Higdon et al. (2007), Hallström and Janke (2008), Chatterjee et al. (2009), and Nyakatura and Bininda-Emonds (2012).

multiple $GPX6_{Cys}$ lineages. We evaluated the impact of natural selection by calculating independent dN/dS ratios (Yang 2007) for each branch of the mammalian tree (Fig. 1b), including ancestral branches. Here, we excluded the Sec-to-Cys site from calculations of the dN/dS ratio to only capture amino acid evolution accompanying the loss of Sec. Indeed, the dN/dS ratios appear larger in $GPX6_{Cys}$ lineages compared with neighboring $GPX6_{Sec}$ lineages, suggesting faster evolution along the branches with Cys (supplementary fig. S2, Supplementary Material online). We tested this hypothesis with a branch model likelihood ratio (LR) test PAML (Yang 2007) and found that contrasting the dN/dS ratios of $GPX6_{Cys}$ lineages in the branches where Sec was lost (Fig. 1b, dashed green branches) to $GPX6_{Cys}$ lineages in the branches inheriting this loss (Fig. 1b, solid green branches) and $GPX6_{Sec}$ lineages (Fig. 1b, solid red branches) does indeed support a higher dN/dS ratio in the branches where Sec was substituted for Cys (LR test; $P = 0.002$; dN/dS = 0.370 dashed green vs. 0.279 solid green vs. 0.217 solid red branches in Fig. 1b). Therefore, significant additional amino acid evolution must have accompanied the loss of Sec.

However, dN/dS inflation across $GPX6_{Cys}$ lineages is still under 1, which does not clearly indicate that positive selection has acted along the branches where Sec was lost, rather than relaxed constraint in $GPX6_{Cys}$. Since positive selection acting on $GPX6_{Cys}$ enzymatic properties would mainly impact the catalytic domain, which is otherwise under strong constraint, it is unsurprising that the dN/dS ratio across the whole gene does not reach 1. We thus separately performed the LR test in its three domains: the N-terminus, the GPX domain, and the C-terminus, as defined in the Pfam database (Mistry et al. 2021). The GPX domain and, to a lesser extent, the C-terminus domain are essential for the activity of the enzyme, with two (U/C, Q) and two (W, N) key catalytic residues in the GPX and C-terminus domains, respectively, making a catalytic tetrad (Toppo et al. 2008; Tosatto et al. 2008; Cheng and Arnér 2017), which we found conserved across $GPX6_{Sec}$ and $GPX6_{Cys}$ lineages. In contrast, the N-terminus is not considered essential for catalysis.

As expected, constraint grows from the N-terminus, to the C-terminus, and to its highest degree in the GPX domain, as indicated by their dN/dS ratios (Table 1). However, the dN/dS ratio of the GPX domain is significantly larger in $GPX6_{Cys}$ lineages at the time Sec was lost when compared with the other branches (LR test; $P = 2 \times 10^{-5}$; dN/dS = 0.384 dashed green vs. 0.186 solid green vs. 0.130 solid red branches in

Table 1
dN/dS ratios across lineages, proteins, and protein domains

			dN/dS in branches where GPX6 has				Convergent sites
Protein	**Region**	**Sec**	**Exchanged Sec for Cys**	**Inherited Cys**	**All**	***P*-value**	**Number**
$GPX6_{Cys}$	**Full length**	0.217	**0.370**	0.279	0.256	**0.002**[a]	**22**
	N-terminus	0.436	0.411	0.671	0.460	0.268	3
	GPX domain	0.130	**0.384**	0.186	0.184	**2×10^{-5}**[b]	**12**
	C-terminus	0.174	0.258	0.250	0.203	0.157	7
$GPX1_{Sec}$	GPX domain	0.064	0.040	0.069	0.060	0.534	0
$GPX2_{Sec}$		0.075	0.042	0.038	0.060	0.191	0
$GPX3_{Sec}$		0.094	0.108	0.056	0.091	0.439	1
$GPX4_{Sec}$		0.062	0.007	0.203	0.061	1×10^{-4}[b]	0
$GPX5_{Sec}$		0.233	0.145	0.219	0.212	0.227	4
$GPX7_{Cys}$	GPX domain	0.083	0.080	0.117	0.088	0.712	0
$GPX8_{Cys}$		0.223	0.155	0.198	0.207	0.616	0

dN/dS ratios in lineages where GPX6 has Sec (Fig. 1b, solid red branches), has gained Cys (Fig. 1b, dashed green branches), or inherited it (Fig. 1b, solid green branches) and the number of identified convergent sites between lineages where GPX6 has gained Cys (Fig. 1b, dashed green branches). dN/dS ratios and number of identified convergent sites for the GPX domain in other GPX proteins. The LR test contrasts one ratio for all branches (null hypothesis) to different ratios among groups of branches. *P*-values are obtained from a χ^2 distribution with df = 2. In bold when significant and accompanied by sites under convergent evolution across $GPX6_{Cys}$ lineages.
[a]$P < 0.005$. [b]$P < 0.0005$.

Fig. 1b), a pattern not observed in the N- and C-terminus. This is again suggestive of an increased level of evolutionary change on the active GPX domain surrounding the time when Sec is abandoned in catalysis.

Localized Evolutionary Signatures to the Sec-to-Cys Exchange

We investigated whether this observation was exclusive to $GPX6_{Cys}$ and therefore indicative of rapid evolution associated with the Sec-to-Cys exchange rather than on the overall antioxidant function of the GPX domain, as others have suggested (Tian et al. 2021). Indeed, when isolating this domain within other enzymes in the GPX family, we found no evidence of dN/dS inflation (Table 1; supplementary table S1, Supplementary Material online) in the lineages where Sec was lost in GPX6 (analogous dashed green branches in Fig. 1b in other GPXs). Within the other enzymes, there is only a significant inflation of this domain in $GPX4_{Sec}$ (Table 1) in the lineages that contain the loss of Sec in GPX6 (analogous solid green branches in Fig. 1b in $GPX4_{Sec}$), unrelated to the Sec-to-Cys exchange. We reason that the dN/dS ratio of the GPX domain in $GPX6_{Cys}$ at the time of Sec loss is unusually large for proteins of the GPX family.

We then asked whether the higher dN/dS is due to positive selection on individual sites that accompany the substitution of Sec for Cys in $GPX6_{Cys}$. We contrasted $GPX6_{Cys}$ lineages for this domain at the time of Sec loss (Fig. 1b, dashed green branches) with all other lineages and found a branch-site model LR test (PAML; Yang 2007) significant (LR test; $P = 0.046$; supplementary table S2, Supplementary Material online). This is suggestive of sites under positive selection in the GPX domain when Sec was lost.

Given the sequence conservation observed in orthologous proteins such as GPX6, we then consider if epistatic and pleiotropic constraints are shared over $GPX6_{Cys}$ lineages. By extension, we consider if the fitness of mutations, which depend on their interaction with the genetic background and their effects on multiple functions, are similar across the orthologous proteins and if the same mutations may be repeatedly under selection. To address this, we ask if many of the same amino acid exchanges are shared amongst the mammalian lineages where Cys is lost. From here on, these are referred to as convergent sites.

To identify these convergent sites, we use a method that uses the alignments of the $GPX6_{Cys}$ proteins with the known mammalian phylogeny to infer ancestral proteins and identify sites which have changed to the same amino acid or repeatedly changed over pairs of $GPX6_{Cys}$ lineages (CONVERG2; Zhang and Kumar 1997). These identified convergent sites are most common between lineages where Sec was lost (Fig. 1b, dashed green branches) (supplementary table S3, Supplementary Material online). We therefore suggest that, following the loss of Sec, positive selection acts mostly on a particular subset of mutations.

Convergence Concentrates in the Catalytic Domain

Most convergent sites are found in the GPX domain and the least in the N-terminus (54.6% in the GPX domain, followed by 31.8% and 13.6% in the C-terminus and N-terminus, respectively). This approximately matches the lengths of each domain (113 sites of the GPX domain compared with the 65 and 39 sites of the C-terminus and N-terminus, respectively), despite the relatively higher constraint expected on the GPX domain. Moreover, convergence is largely subdued

in the GPX6$_{Cys}$ lineages inheriting the loss of Sec (Fig. 1b, solid green branches) and minimal in the GPX6$_{Sec}$ lineages and other GPX$_{Sec}$ and GPX$_{Cys}$ enzymes (supplementary tables S3 to S10 and fig. S3, Supplementary Material online). Further, simulations of protein evolution incorporating the accelerated rate of amino acid change in GPX6$_{Cys}$ sequences, including the GPX domain, cannot reproduce the pattern of convergence observed between these lineages at the time of Sec loss (Seq-gen (Rambaut and Grassly 1997); supplementary figs. S7 to S8, Supplementary Material online). However, these simulations show that the few weak convergence signatures in other GPXs are as expected, based on their respective rate of amino acid change (supplementary figs. S9 to S15, Supplementary Material online), and we presume they result from chance. We therefore infer that signatures of convergence are exclusive to GPX6$_{Cys}$ and strongest in its GPX domain at the time that Sec was lost. This suggests that GPX6$_{Cys}$ lineages have similar evolutionary trajectories following the Sec-to-Cys exchange and a narrow path in which to restore fitness, which may be in part due to few epistatically relevant differences in the genetic backgrounds of GPX6 proteins amongst placental mammals.

Strong Convergence between Most Related GPX6 Lineages with Cys

While convergent sites are identified amongst all GPX6$_{Cys}$ lineages (supplementary table S3, Supplementary Material online), we inferred the highest level of convergence between the basal *Eumuroida* (Fig. 1b, dashed green line in box) and its genetically closer GPX6$_{Cys}$ lineages, particularly the rabbit lineage (supplementary fig. S3 and table S3, Supplementary Material online). To maximize our power to identify signatures of convergence amongst GPX6$_{Cys}$ lineages, we therefore focus on the sites that change along the root of *Eumuroida*. We find 25 sites at the root of the *Eumuroida* (Fig. 2a, dashed green branch) that changed alongside Sec and found that 14 of them bear signatures of convergence across the GPX6$_{Cys}$ lineages (Fig. 2a, green box). These convergent sites are, again, mainly in the GPX catalytic domain, 64.3% of them (supplementary table S2, Supplementary Material online), and are enriched for the signatures of positive selection that we observe along branches leading to *Eumuroida*, rabbit, and marmoset–squirrel monkey (Mann–Whitney *U* test, $P = 1.573e^{-7}$; Yang et al. 2005). We also find signatures of positive selection, albeit weaker, in convergent sites in the GPX6$_{Cys}$ lineages with the loss of Sec (Mann–Whitney *U* test, $P = 0.007$) (Fig. 2a, solid green branches) but not preceding it, in agreement with adaptive convergence acting on the Sec-to-Cys exchange.

Because adaptive convergence can mimic shared ancestry (Edwards 2009), it may distort the topology of the species phylogeny (Fig. 1b), and indeed, we found that a tree reconstructed (PhyML; Guindon et al. 2010) from the GPX domain pulls the rabbit lineage closer to the *Eumuroida* clade (supplementary fig. S4d, Supplementary Material online). This, to a lesser extent, is also observed with the C-terminus (supplementary fig. S4e, Supplementary Material online), which contributes to catalysis, but is not observed with the N-terminus domain (supplementary fig. S4c, Supplementary Material online) nor with other GPX proteins (supplementary fig. S5, Supplementary Material online). The strongest departure from the species tree is reconstructed from the 14 convergent sites changing at the root of the *Eumuroida* 23 to 26 Mya (Huchon et al. 2002), plus the Sec-to-Cys substitution (Fig. 2b). In this tree, despite their large divergence, the GPX6$_{Cys}$ species form two clades. One clade shows the rabbit lineage sharing a most recent common ancestor, to the exclusion of all other species, with the 64 million years apart *Eumuroida*, and the other clade groups the remaining 100 million years apart from GPX6$_{Cys}$ lineages (Hallström and Janke 2008). This supports a role for convergence (Edwards 2009) in driving adaptive changes, perhaps compensating for loss of enzymatic activity, or opening new properties, as it seems to have happened with GPX5$_{Cys}$, GPX7$_{Cys}$, and GPX8$_{Cys}$ enzymes that lost Sec much earlier (supplementary fig. S1a, Supplementary Material online) (Herbette et al. 2007; Chen et al. 2016).

Loss of Activity in Reconstructed Ancient GPX6 Proteins with Cys

The catalytic activity of enzymes that have exchanged Sec to Cys has been shown to decrease (Gromer et al. 2003). We thus expect to recapitulate this here: functional catalytic activity of Eu-GPX6$_{Sec}$ and a nonfunctional or reduced catalytic activity of Eu-GPX6$_{Cys}$. Given the signatures of adaptive convergence in *Eumuroida* following the loss of Sec, we also hypothesize the same compensatory mutations have been under natural selection across multiple mammalian lineages losing Sec to recover catalytic function in GPX6$_{Cys+25}$.

To evaluate this, we reconstructed three ancient proteins inferred to exist 23 to 26 Mya at the root of this clade (Fig. 2a, dashed green branch) and assessed their catalytic activity experimentally and computationally. These proteins are (i) the ancestral protein before the loss of Sec, Eu-GPX6$_{Sec}$, taken from the common ancestor of the *Eumuroida* and Jerboa species 26 Mya (Huchon et al. 2002); (ii) the same ancestral protein with Cys instead of Sec, Eu-GPX6$_{Cys}$; and (iii) the ancestral but later-day protein with Cys and 25 other amino acids changes, Eu-GPX6$_{Cys+25}$, taken from the common ancestor of the *Eumuroida* species 23 Mya (15 of 26 these amino acid changing sites, including the Cys site, have signatures of adaptive convergence; Fig. 2a). A further modern protein and 2 synthetic variants were also assessed: the modern

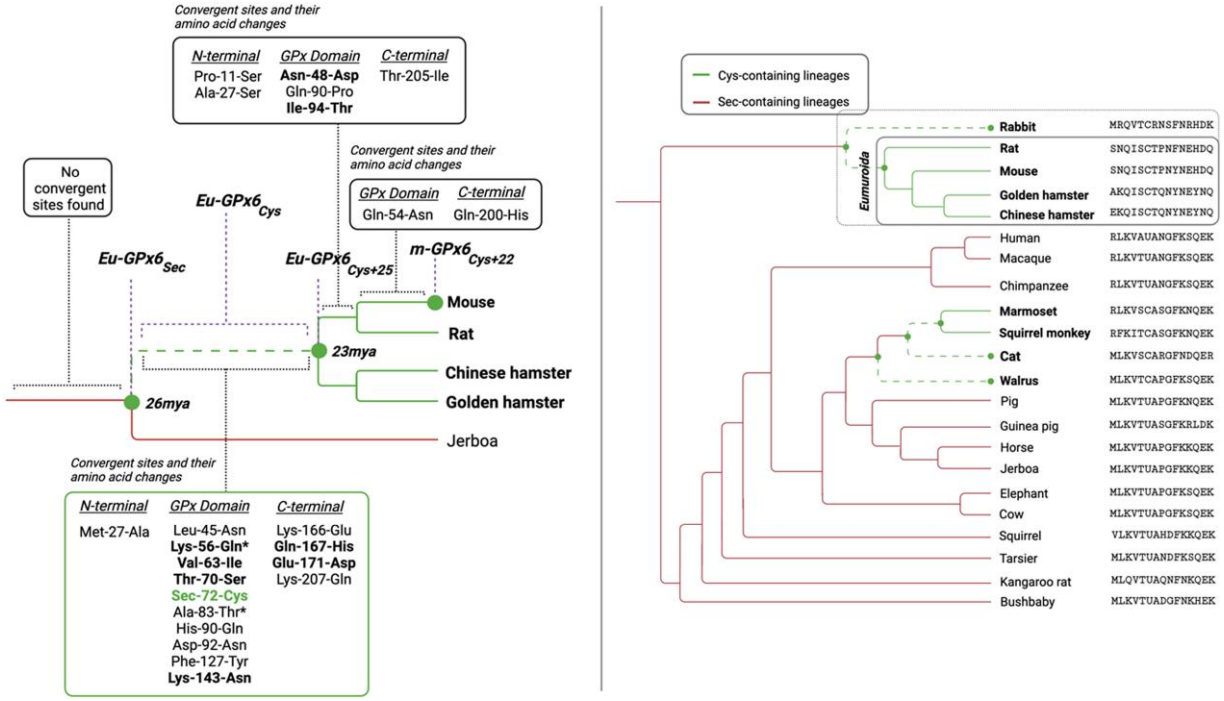

Fig. 2.—Convergence in the GPX6 phylogeny. a) Topology of the phylogeny of the *Eumuroida* GPX6$_{Cys}$ clade, green branches, with the Jerboa GPX6$_{Sec}$ lineage, red branch, as an outgroup. The basal *Eumuroida* lineage, dashed green branch, abandoned Sec in catalysis for Cys with a burst of 25 additional amino acid changes 23 to 26 Mya. Fourteen of these 25 changing amino acid sites, plus the Sec-to-Cys site, have signatures of convergence (CONVERG2; Zhang and Kumar 1997) across GPX6$_{Cys}$ lineages (Fig. 1b, green branches). Sites that have repeatedly changed in the GPX6$_{Cys}$ lineages toward similar or the same amino acid are shown in bold (green box). No convergent sites are found immediately before the loss of Sec and only few immediately after. Further, the * denotes sites with a posterior probability of positive selection in the upper 90th percentile across the GPX domain in GPX6$_{Cys}$ lineages, which are significantly enriched at the time Sec was abandoned. b) Topology of the phylogenetic tree, with midpoint rooting, from the 14 convergent sites accompanying the Sec-to-Cys substitution (sites corresponding to the positions given in the large green box in Fig. 2a, shown to the left of the phylogeny) in the basal *Eumuroida* GPX6$_{Cys}$ lineage (Fig. 2a, dashed green branch). In sharp contrast to the species phylogeny (Fig. 1b), the GPX6$_{Sec}$ lineages now form two clades.

mouse protein, m-GPX6$_{Cys+22}$, with 22 additional amino acid changes (19 substitutions and a 3 C-terminal extension) from Eu-GPX6$_{Cys+25}$ and no clear signatures of adaptive convergence (Fig. 2a), the modern mouse protein with Cys mutated to Sec (m-GPX6$_{Sec+22}$), and the modern mouse protein with Cys mutated to redox inactive serine (Ser; m-GPX6$_{Ser+22}$) for comparisons of activity with the Sec and Cys variants.

We reconstructed these ancient and modern proteins and produced them as recombinant proteins heterologously expressed in *Escherichia coli*. The Sec insertion system in bacteria is noncompatible with mammalian selenoprotein-encoding genes, hampering the production of proteins with Sec; we thus employed a method we recently developed utilizing UAG redefined as a Sec codon in a release factor 1–deficient *E. coli* host strain lacking other UAG codons (Cheng and Arnér 2017). We first compared the catalytic activity of Eu-GPX6$_{Sec}$ and Eu-GPX6$_{Cys}$ with H_2O_2 as the peroxide substrate and GSH as the reducing agent, with the expectation that substitution of Sec for Cys would lower its turnover. Indeed, the ancient Eu-GPX6$_{Sec}$ protein displays the classic peroxidase activity of Sec-containing GPX enzymes, whereas Eu-GPX6$_{Cys}$ had almost no activity for this reaction (Fig. 3a).

We therefore observe the hypothesized a large drop in catalysis from Eu-GPX6$_{Sec}$ to Eu-GPX6$_{Cys}$. Along the same basal *Eumuroida* lineage, we observe signatures of adaptive evolution (Fig. 2a), which may be assumed to then recover such a lost activity. However, when measuring the catalytic activity in Eu-GPX6$_{Cys+25}$ on H_2O_2 with GSH, we do not observe the classic GPX activity to be recovered. Hence, the 25 additional changes along the basal *Eumuroida* lineage, inferred to be under adaptive convergence, do not appear to recover the original enzymatic function.

Classic GPX Activity Unobserved in the Modern Mouse Protein

We then evaluated the activity of the extant m-GPX6$_{Cys+22}$ protein (Fig. 2a), which is 90% identical to Eu-GPX6$_{Cys+25}$. This protein is expressed in the mouse embryo, testis, olfactory epithelium, and brain (Kryukov et al. 2003; Shema et al. 2015; Goltyaev et al. 2020) with clear function

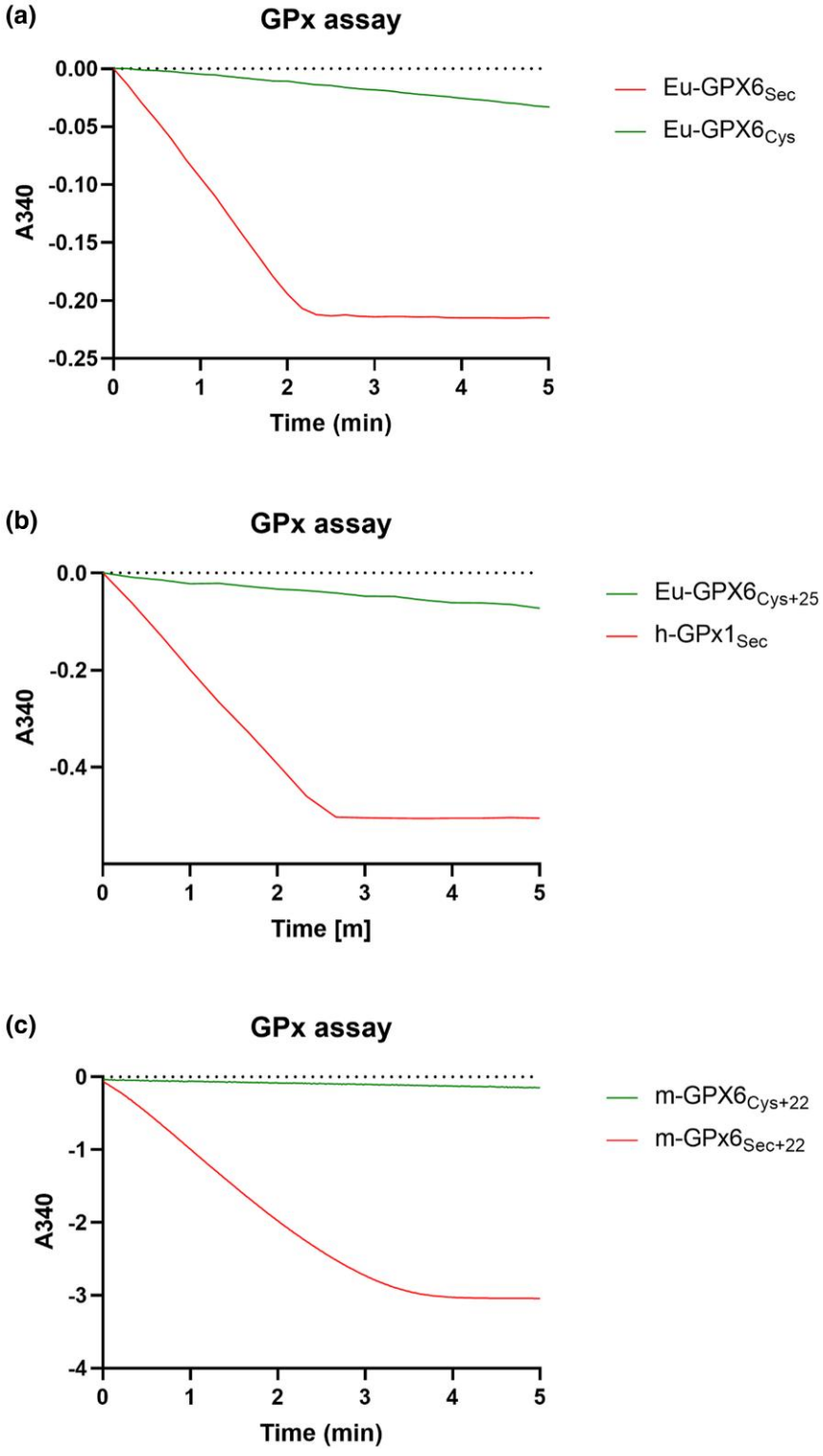

Fig. 3.—Experimental assessment of catalytic ability in ancestral and modern GPX6 proteins. a) Experimental assessment of peroxidase reaction with H_2O_2 as a substrate for ancient Eu-GPX6$_{Sec}$ (red) and Eu-GPX6$_{Cys}$ (green). NADPH consumption by GR is indicated by the decrease in absorbance at 340 nm over time in the coupled assay (see Materials and Methods for further details). b) Equivalent assay for ancient Eu-GPX6$_{Cys+25}$ (green), which has very limited activity compared with human GPx1 (red) used here as a positive control. c) Equivalent assay for modern m-GPX6$_{Cys+22}$ (green), again with scant activity, which is recovered once this protein is mutated to contain Sec, m-GPX6$_{Sec+22}$ (red).

(knocking down in modern mice results in neurological consequences (Shema et al. 2015)). However, this Cys-containing variant also lacks classic GPX activity with H_2O_2 and GSH (Fig. 3c), with GPX activity also unobserved when testing the additional substrate cumene hydroperoxide (supplementary fig. S15, Supplementary Material online). This suggests that its modern function lies outside of that of a classic GPX protein.

It can be suggested that increased expression of the Cys-containing proteins could recover catalytic activity to a functional level. However, the activity of the extant m-GPX6$_{Cys+22}$ protein displays no detectable peroxidase activity (activity below −0.02A340; supplementary fig. S16, Supplementary Material online), almost identical to that of the redox inactive synthetic m-GPX6$_{Ser+22}$ protein. Increasing the expression of the Cys protein, even by a factor of ten or higher, is therefore unlikely to result in a meaningful enzymatic activity and would not be expected to recover classic GPX function. Hence, we suggest the observed adaptive changes in the evolution of this protein do not simply act to recapitulate Sec activity. Instead, we propose that the signatures of adaptive convergence represent the convergent evolution of a new, as of yet, underdetermined function, as observed in other GPX proteins that have lost Sec (Herbette et al. 2007; Chen et al. 2016).

Binding of GSH Unaffected by Acquisition of Cys

Interestingly, classic GPX activity was reacquired when Cys was mutated back into Sec, producing the synthetic m-GPX6$_{Sec+22}$ variant (Fig. 3c). We therefore asked if classic GPX activity was likely to be recovered with the replacement of Cys with Sec in all ancestral Cys proteins. To assess this, we computationally assessed the binding of GSH to the Eu-GPX6$_{Cys}$, Eu-GPX6$_{Cys+25}$, and m-GPX6$_{Cys+22}$ proteins by running protein–ligand binding energy landscape explorations via the PELE software (Borrelli et al 2005), using GSH and GSH disulfide as ligands.

Models for all proteins were obtained using the AlphaFold2 algorithm (Jumper et al. 2021) and showed no significant structural changes in their overall fold shape when compared among them (largest alpha-carbon RMSD: 2.1 Å over 218 residues). The reactive free energy minimum was inferred by plotting the free energy of the sampled trajectories along the slowest kinetic coordinate computed for all our Monte Carlo simulations (IC1) and the reactive distance between the sulfur atoms of GSH and the catalytic Cys (Fig. 4a). These simulations used a protocol that first diversifies the possible GSH binding modes around the catalytic residues to then further explore the lowest reactive energy configurations uncovered by this first broad sampling. This adaptive sampling technique ensures more convergent results despite the initial ligand starting position (Gilabert et al. 2019).

The energy landscapes inferred for Eu-GPX6$_{Cys}$, Eu-GPX6$_{Cys+25}$, and m-GPX6$_{Cys+22}$ all show similar minima at reactive distances (a vertical dashed line is used as threshold at 4 Å). Since the IC1 composite coordinate is constructed from a common reference for the ligand Cartesian coordinates for all simulations (see Materials and Methods), we infer that all three variants have similar binding modes for GSH. All variants also show similar values at their binding energy minima in the reactive region (supplementary fig. S17, Supplementary Material online), thereby indicating they interact with similar strengths with GSH.

Our computational analysis thus suggests that the binding of GSH and overall structures of the enzymes have not been adversely affected by the acquisition of Cys or other mutations. Further, the convergent amino acid substitutions, possibly facilitating a new function of the protein, are mainly located on the enzyme's surface (Fig. 4b). Given that this is the case for GPX6 in other lineages losing Sec (supplementary fig. S17, Supplementary Material online), we again suggest that a new function is shared across all mammalian GPX6 proteins having lost Sec.

Discussion

Substituting Sec for Cys in GPX6, and thereby abandoning selenium for sulfur, leads to a burst of evolutionary activity in lineages sharing this exchange. This may be an example of an "evolutionary Stokes shift," where the evolutionary activity of a protein increases following a significant amino acid exchange, and new mutations may gradually increase the fitness of the original substitution over time (Pollock et al. 2012). The amino acid changes are not only concentrated in the functional domain but are often shared across GPX6$_{Cys}$ lineages, suggesting a narrow evolutionary range for GPX6 to maintain or increase fitness when losing Sec. As supported by increased convergence between the closest related lineages, this is likely a path under both epistatic and pleiotropic constraints; the fitness of mutations is dependent on interactions with other conserved catalytic sites and preservation of other key enzymatic properties (Sharir-Ivry and Xia 2021).

We initially hypothesized that the observed signatures of adaptive convergence represented a recovery of catalytic activity across GPX6$_{Cys}$ lineages. Indeed, thioredoxin reductase in *Drosophila melanogaster* has been shown to recover up to 50% of its catalytic activity with compensatory mutations following its loss of Sec (Gromer et al. 2003), and we viewed it possible that compensatory mutations also acted to restore catalytic activity here. However, classic GPX activity is not recovered by the amino acid changes along the *Eumuroida* lineage, nor is it likely to be compensated by the higher expression of Cys when the activity is undetectable. Still, contemporary GPX6$_{Cys}$ proteins are functional,

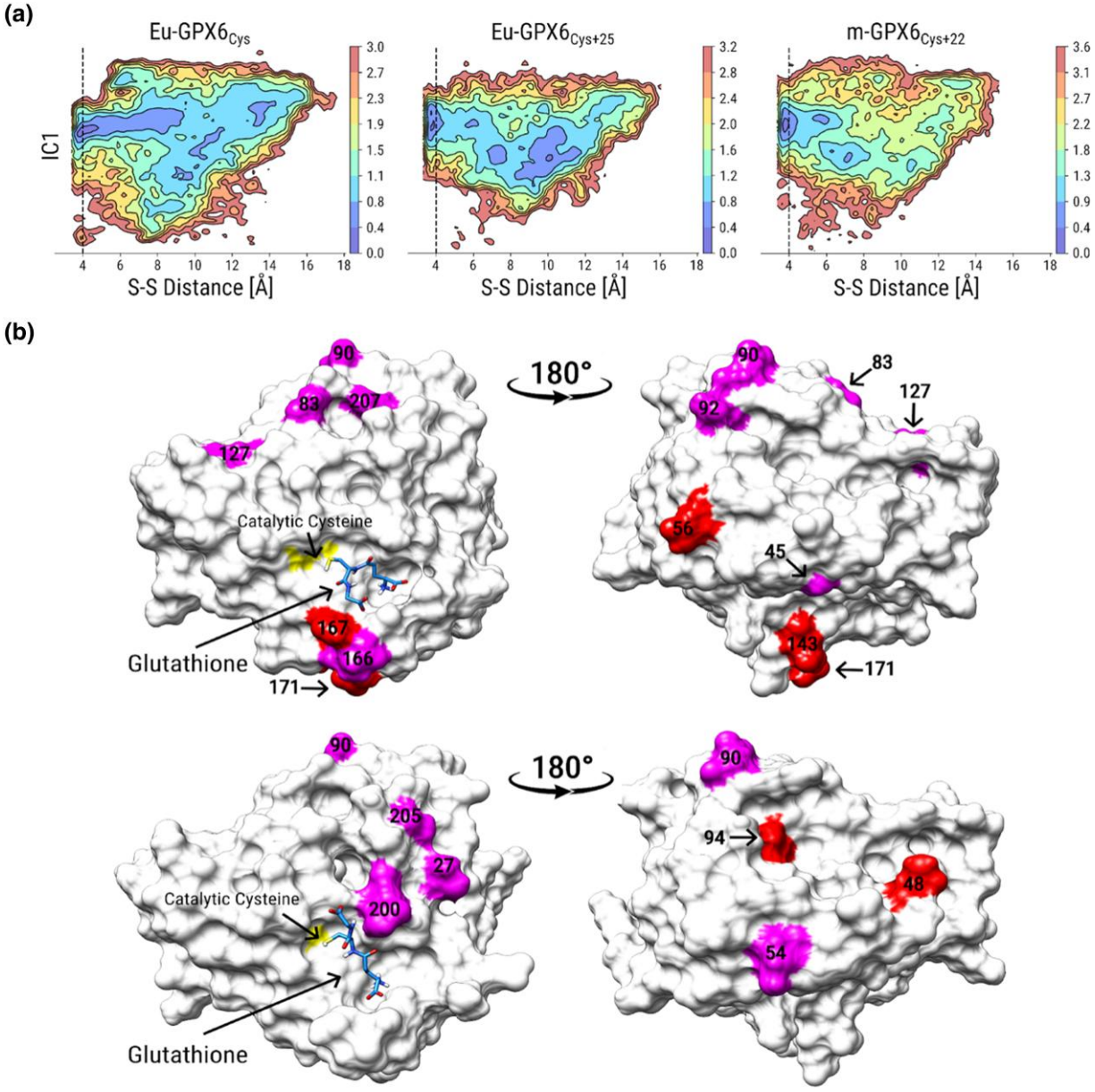

Fig. 4.—Computational analysis of enzyme structure and catalytic reaction. a) Free energy profiles for the docking of GSH to Eu-GPX6$_{Cys}$ (left), Eu-GPX6$_{Cys+25}$ (center), and m-GPX6$_{Cys+22}$ (right). The *x* axis represents the distance between the catalytic Cys sulfur atom and the ligand's sulfur atom, while the *y* axis shows the slowest TICA coordinate (IC1). The vertical dashed line represents a 4 Å distance, with the free energy minimum in the three enzymes within this reactive catalytic distance. b) Convergence patterns (Fig. 2a) from Eu-GPX6$_{Cys}$ to Eu-GPX6$_{Cys+25}$ (top) and from Eu-GPX6$_{Cys+25}$ to m-GPX6$_{Cys+22}$ (mouse-GPX6) (bottom). Sites converging toward similar (magenta) or the same (red) amino acids are shown with their sequence position. The catalytic Cys (yellow) is shown with the GSH best binding energy conformation (green) sampled during docking simulations.

expressed in the mouse embryo, testis, olfactory epithelium, and brain (Kryukov et al. 2003; Shema et al. 2015; Goltyaev et al. 2020), and when knocked down result in a deleterious neurological phenotype (Shema et al 2015). The observed signatures of adaptive convergence across GPX6$_{Cys}$ proteins therefore do not represent a recovery of classic GPX function but a shared evolutionary trajectory requiring a more considered explanation.

Losing Sec, and with it classic GPX function, may have widened the number of tolerable mutations in the highly similar functional domain of GPX6$_{Cys}$ proteins, resulting in the observed increase of evolutionary rate and similarly mutated sites across mammalian lineages. The contemporary function of GPX6$_{Cys}$ proteins, in this scenario, would therefore not be novel and would instead represent a maintained ancestral function of which drives the still-narrow range of

tolerable mutations. The apparent tolerable loss of Sec itself remains a question, but it is of course possible that the activity of other Sec-containing GPX proteins compensate for the loss of classic GPX activity to the extent that catalytic function of the protein may be lost.

Alternatively, the observed signatures of adaptive convergence may represent shared evolutionary trajectories toward novel properties or novel substrates. Indeed, other Cys-containing GPX proteins act on alternative substrates for peroxidation (Nguyen et al. 2011; Taylor et al. 2013; Buday and Conrad, 2021), and it is possible that $GPX6_{Cys}$ has also evolved a preference for a novel substrate, one differing from hydrogen peroxide and cumene hydroperoxide tested here, for this ultimate function. Thus, perhaps similar to old and single losses of Sec occurring hundreds of millions of years ago in $GPX5_{Cys}$, $GPX7_{Cys}$, and $GPX8_{Cys}$ enzymes (Trenz et al. 2021), we suggest that $GPX6_{Cys}$ proteins may have also gained yet unidentified abilities, though acquired more recently, independently, and convergently across lineages, instead of just recovering the catalytic rate of their previous reaction. Further, $GPX6_{Cys}$ proteins across mammals appear to be able to recover their classic GPX function with Sec, which we view as in agreement with the loss of Sec resulting in subtly different, but related, catalytic properties in an enzyme now apparently devoid of its classic function. Again, we propose it likely that the activity of other Sec-containing GPX proteins compensated for the immediate loss of classic GPX activity when Sec was lost in GPX6, but here, this allowed mutations to accumulate along $GPX6_{Cys}$ lineages and novel properties to develop. Only comprehensive functional characterizations of these individual $GPX6_{Cys}$ enzymes, and their interactions with other GPX proteins, in mammals will provide insights into their current functional roles, be that those related to peroxidation or entirely novel or undocumented.

In conclusion, we present the first evidence for molecular convergence of changes in proteins when abandoning unusual selenium in catalysis for common sulfur, hence ablating activity. These concerted changes follow a certain path, maintaining some enzymatic properties and possibly adding new ones. Because multiple nonvertebrate species have completely abandoned enzymatic selenium for sulfur, we wonder whether other convergent adaptations leading to uncharted functions remain hidden in nature.

Materials and Methods

The GPX6 coding sequences and proteins (alongside those for other members of the GPX family) for 22 present-day mammal species were taken from SelenoDB 2.0 (Romagné et al. 2014) or Ensembl (Yates et al. 2020). Of these species, nine contain Cys instead of Sec as their catalytic residue of GPX6 (Fig. 1). We aligned the coding sequences using MAFFT (Katoh et al. 2019), and uncertain positions with an average posterior probability below 0.95, as calculated by HMMER (Durbin et al. 1998; Potter et al. 2018), were removed from further analysis.

We used PAML (Yang 2007) to reconstruct ancestral nodes across the mammalian GPX6 tree and to infer independent losses of Sec. Here, the accuracy of ancestral node reconstruction for all but one node is estimated to be above 96% (with the outlier node being the most basal, with an accuracy of 88.45%; see supplementary fig. S1, Supplementary Material online). The ancestral sequences were also then corroborated with Ancestor v1.1 (Diallo et al. 2010) and FastML (Moshe and Pupko 2019) (see Supplementary Material online). Given this multiple-pronged approach in reconstructing ancestral sequences and the well-resolved mammalian phylogeny, the ancestral sequences are reported with confidence. Independent dN/dS ratios for each mammalian was computed using the free-ratio model of the CODEML package from PAML (Yang 2007), whereas the branch model was used to explicitly test if an increased rate of evolution occurred in specific groups of lineages. We used the latter to compare dN/dS between branches with Sec (Fig. 1; solid red branches), branches where Sec is exchanged for Cys (Fig. 1, dashed green branches) and branches where Cys is maintained (Fig. 1, solid green branches). This was repeated for the three separate domains of the protein: N-terminus, GPX domain, and C-terminus. We computed the LR of our branch model to the null model (which assumes a singular dN/dS value across branches) and used this to calculate the significance of difference in fit between the two models (Table 1). This was repeated for other genes in the GPX family, testing for differences in dN/dS across analogous groups of lineages. We tested for selection acting on individual sites across specific branches using the branch-site model in PAML (Yang 2007) both across the entire protein and on the catalytic GPX domain.

Based on the alignments of the $GPX6_{Cys}$ proteins and mammalian phylogeny, we then consider each pair of $GPX6_{Cys}$ lineages and identify sites that have changed from the common ancestor along both lineages, now termed convergent sites (supplementary table S3, Supplementary Material online). This was done using the CONVERG2 program (Zhang and Kumar 1997), which infers ancestral proteins along the mammalian phylogeny to count the convergent changes for each pair of lineages (as well as calculating the expected frequencies of convergent changes between pairs of lineage). We further compared these expected frequencies to the observed frequencies of convergent changes for each pair of $GPX6_{Cys}$ lineages.

Where the sequences for the species containing Cys in GPX6 were available, equivalent analyses were run on other members of the GPX family. We tested for enrichment of signatures of positive selection, as calculated from the branch-site model in PAML along the branch leading to

squirrel monkey–marmoset, the Eumuroida branch, and the branch leading to rabbit, across these convergent sites using a Mann–Whitney *U* test. We also tested enrichment of selection signatures in sites inferred to have changed over the *Eumuroida* branch and with signatures of convergence in $GPX6_{Cys}$ lineages (Fig. 2a). PHYML was used to reconstruct the mammalian tree according to the 14 identified sites (Guindon et al. 2010) (Fig. 2b).

The evolution of the GPX6 protein sequence was simulated using Seq-Gen (Rambaut and Grassly 1997), using our inferred mammalian ancestral sequence, the JTT model of amino acid substitution, and tree lengths given by the rate of amino acid changes (taken as the dN value calculated from the CODEML package in PAML (Yang 2007). The distribution of convergent changes, as calculated from CONVERG2 (Zhang and Kumar 1997), from 1,000 simulation runs, was then plotted and compared with the observed number of convergent site changes (supplementary fig. S6, Supplementary Material online). The equivalent methodology was run for the catalytic GPX domain, as well as for other GPX proteins.

We reconstructed the Eu-$GPX6_{Sec}$, Eu-$GPX6_{Cys}$, and Eu-$GPX6_{Cys+25}$ proteins (Fig. 2a), as inferred from PAML (Tian et al. 2021), from heterologous expression in *E. coli*. The catalytic activities of these proteins, and the modern mouse protein, were evaluated by measuring the peroxidation activity on H_2O_2. The sequences of the reconstructed ancient proteins and recombinant proteins, along with purification and assay protocols, are given in Supplementary Material online.

Using AlphaFold2 (Jumper et al. 2021), we built structures for the GPX6 orthologs and ancestral sequences to run protein–ligand binding energy landscape explorations using the PELE software (Borrelli et al. 2005), using GSH and GSH disulfide as ligands. Structures were first prepared with the Protein Preparation Wizard of the Schrodinger suite (Sastry et al. 2013) by setting the protonation state at pH 7. Initial conformations were computed with the GLIDE algorithm (Friesner et al. 2006), and the best-scoring conformations were selected among conformations with a reactive distance, between the ligand and the catalytic Cys sulfur atoms (S-S distance), lower than 4 Å. From these starting conformations, a first PELE simulation was run to identify catalytic poses with low global energies and reactive distances below a 4 Å threshold. This first simulation equilibrates the system without ligand constraints to obtain a diversified set of low-energy configurations. We used the lowest binding energy poses within the S-S distance threshold to run a second PELE simulation, focusing on the sampling on this reactive region. All obtained simulation trajectories were aligned to a common coordinate frame of reference by aligning each frame to the same protein structure. A time-structure independent component analysis (TICA) was performed to find the common slowest-relaxing feature combination (Molgedey and Schuster 1994) with the PyEMMA library (Scherer et al. 2015) by using as features only the ligand Cartesian coordinates. The probabilities of visiting the slowest TICA coordinate (IC1) according to the S-S distance were plotted as free energy maps for each simulation (Fig. 4 and supplementary fig. S17, Supplementary Material online).

Supplementary Material

Supplementary material is available at *Genome Biology and Evolution* online.

Acknowledgments

Figures 1 and 2 and supplementary figs. S1 and S3 and S7 to S15, Supplementary Material online, were created with the use of BioRender.

Funding

E.S.J.A. and Q.C. are funded by the Karolinska Institutet, the Knut and Alice Wallenberg Foundations (KAW 2019.0059), the Swedish Cancer Society (21 1463 Pj), the Swedish Research Council (2021-02214), the National Laboratories Excellence program under the National Tumor Biology Laboratory project (2022-2.1.1-NL-2022-00010), and the Hungarian Thematic Excellence Programme (TKP2021-EGA-44) and the National Research, Development and Innovation Office (NKFIH) grant ED_18-1-2019-0025. M.F., B.O.M., and J.V.-F. are funded by project BIO2017-83650-P, financed by the Spanish Ministerio de Ciencia e Innovación (MCIN). A.M.A. is funded by UCL's Wellcome Institutional Strategic Support Fund 3 (grant reference 204841/Z/16/Z). S.C. and J.R. are funded by NIHR GOSH BRC. The views expressed are those of the authors and not necessarily reflect those of the funding body, including those of the NHS, the NIHR, or the Department of Health.

Data Availability

The data underlying this article are available in the article and in its online Supplementary Material online.

Literature Cited

Arnér ESJ. Selenoproteins—what unique properties can arise with selenocysteine in place of cysteine? Exp Cell Res. 2010:316(8): 1296–1303. https://doi.org/10.1016/j.yexcr.2010.02.032.

Axley MJ, Böck A, Stadtman TC. Catalytic properties of an *Escherichia coli* formate dehydrogenase mutant in which sulfur replaces selenium. Proc Natl Acad Sci U S A. 1991:88(19):8450–8454. https://doi.org/10.1073/pnas.88.19.8450.

Berry MJ, Maia AL, Kieffer JD, Harney JW, Larsen PR. Substitution of cysteine for selenocysteine in type I iodothyronine deiodinase reduces the catalytic efficiency of the protein but enhances its

translation. Endocrinology 1992:131(4):1848–1852. https://doi.org/10.1210/endo.131.4.1396330.
Borrelli KW, Vitalis A, Alcantara R, Guallar V. PELE: protein energy landscape exploration. A novel Monte Carlo based technique. J Chem Theory Comput. 2005:1(6):1304–1311. https://doi.org/10.1021/ct0501811.
Buday K, Conrad M. Emerging roles for non-selenium containing ER-resident glutathione peroxidases in cell signaling and disease. Biol Chem. 2021:402(3):271–287. https://doi.org/10.1515/hsz-2020-0286.
Carter P, Wells JA. Dissecting the catalytic triad of a serine protease. Nature 1988:332(6164):564–568. https://doi.org/10.1038/332564a0.
Castellano S, Andrés AM, Bosch E, Bayes M, Guigó R, Clark AG. Low exchangeability of selenocysteine, the 21st amino acid, in vertebrate proteins. Mol Biol Evol. 2009:26(9):2031–2040. https://doi.org/10.1093/molbev/msp109.
Castellano S, Lobanov AV, Chapple C, Novoselov SV, Albrecht M, Hua D, Lescure A, Lengauer T, Krol A, Gladyshev VN, et al. Diversity and functional plasticity of eukaryotic selenoproteins: identification and characterization of the SelJ family. Proc Natl Acad Sci U S A. 2005:102(45):16188–16193. https://doi.org/10.1073/pnas.0505146102.
Chatterjee HJ, Ho SY, Barnes I, Groves C. Estimating the phylogeny and divergence times of primates using a supermatrix approach. BMC Evol Biol. 2009:9(1):259. https://doi.org/10.1186/1471-2148-9-259.
Chen Y-I, Wei P-C, Hsu J-L, Su F-Y, Lee W-H. NPGPx (GPx7): a novel oxidative stress sensor/transmitter with multiple roles in redox homeostasis. Am J Transl Res. 2016:8(4):1626–1640. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4859894/pdf/ajtr0008-1626.pdf.
Cheng Q, Arnér ESJ. Selenocysteine insertion at a predefined UAG codon in a release factor 1 (RF1)-depleted *Escherichia coli* host strain bypasses species barriers in recombinant selenoprotein translation. J Biol Chem. 2017:292(13):5476–5487. https://doi.org/10.1074/jbc.M117.776310.
Diallo AB, Makarenkov V, Blanchette M. Ancestors 1.0: a web server for ancestral sequence reconstruction. Bioinformatics. 2010:26(1):130–131. https://doi.org/10.1093/bioinformatics/btp600.
Durbin R, Eddy SR, Krogh A, Mitchison G. Biological sequence analysis: probabilistic models of proteins and nucleic acids. 1st ed. Cambridge: Cambridge University Press; 1998. Available from: https://www.cambridge.org/core/product/identifier/9780511790492/type/book.
Edwards SV. Natural selection and phylogenetic analysis. Proc Natl Acad Sci U S A. 2009:106(22):8799–8800. https://doi.org/10.1073/pnas.0904103106.
Friesner RA, Murphy RB, Repasky MP, Frye LL, Greenwood JR, Halgren TA, Sanschagrin PC, Mainz DT. Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. J Med Chem. 2006:49(21):6177–6196. https://doi.org/10.1021/jm051256o.
Gilabert JF, Grebner C, Soler D, Lecina D, Municoy M, Gracia Carmona O, Soliva R, Packer MJ, Hughes SJ, Tyrchan C, et al. PELE-MSM: a Monte Carlo based protocol for the estimation of absolute binding free energies. J Chem Theory Comput. 2019:15(11):6243–6625. https://doi.org/10.1021/acs.jctc.9b00753.
Goltyaev MV, Mal'tseva VN, Varlamova EG. Expression of ER-resident selenoproteins and activation of cancer cells apoptosis mechanisms under ER-stress conditions caused by methylseleninic acid. Gene 2020:755:144884. https://doi.org/10.1016/j.gene.2020.144884.
Gromer S, Johansson L, Bauer H, Arscott LD, Rauch S, Ballou DP, Williams CH Jr, Schirmer RH, Arnér ESJ. Active sites of thioredoxin reductases: why selenoproteins? Proc Natl Acad Sci U S A. 2003:100(22):12618–12623. https://doi.org/10.1073/pnas.2134510100.
Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol. 2010:59(3):307–321. https://doi.org/10.1093/sysbio/syq010.
Hallström BM, Janke A. Resolution among major placental mammal interordinal relationships with genome data imply that speciation influenced their earliest radiations. BMC Evol Biol. 2008:8(1):162. https://doi.org/10.1186/1471-2148-8-162.
Hedges SB. The origin and evolution of model organisms. Nat Rev Genet. 2002:3(11):838–849. https://doi.org/10.1038/nrg929.
Herbette S, Roeckel-Drevet P, Drevet JR. Seleno-independent glutathione peroxidases. More than simple antioxidant scavengers. FEBS J. 2007:274(9):2163–2180. https://doi.org/10.1111/j.1742-4658.2007.05774.x.
Higdon JW, Bininda-Emonds OR, Beck RM, Ferguson SH. Phylogeny and divergence of the pinnipeds (Carnivora: Mammalia) assessed using a multigene dataset. BMC Evol Biol. 2007:7(1):216. https://doi.org/10.1186/1471-2148-7-216.
Huchon D, Madsen O, Sibbald MJJB, Ament K, Stanhope MJ, Catzeflis F, de Jong WW, Douzery EJP. Rodent phylogeny and a timescale for the evolution of Glires: evidence from an extensive taxon sampling using three nuclear genes. Mol Biol Evol. 2002:19(7):1053–1065. https://doi.org/10.1093/oxfordjournals.molbev.a004164.
Jayaraman V, Toledo-Patiño S, Noda-García L, Laurino P. Mechanisms of protein evolution. Protein Sci. 2022:31(7):e4362. https://doi.org/10.1002/pro.4362.
Jensen RA. Enzyme recruitment in evolution of new function. Annu Rev Microbiol. 1976:30(1):409–425. https://doi.org/10.1146/annurev.mi.30.100176.002205.
Johansson L, Gafvelin G, Arnér ESJ. Selenocysteine in proteins—properties and biotechnological use. Biochim Biophys Acta. 2005:1726(1):1–13. https://doi.org/10.1016/j.bbagen.2005.05.010.
Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, et al. Highly accurate protein structure prediction with AlphaFold. Nature 2021:596(7873):583–589. https://doi.org/10.1038/s41586-021-03819-2.
Katoh K, Rozewicki J, Yamada KD. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Brief Bioinform. 2019:20(4):1160–1166. https://doi.org/10.1093/bib/bbx108.
Kim M-J, Lee BC, Hwang KY, Gladyshev VN, Kim H-Y. Selenium utilization in thioredoxin and catalytic advantage provided by selenocysteine. Biochem Biophys Res Commun. 2015:461(4):648–652. https://doi.org/10.1016/j.bbrc.2015.04.082.
Kryukov GV, Castellano S, Novoselov SV, Lobanov AV, Zehtab O, Guigó R, Gladyshev VN. Characterization of mammalian selenoproteomes. Science 2003:300(5624):1439–1443. https://doi.org/10.1126/science.1083516.
Lee SR, Bar-Noy S, Kwon J, Levine RL, Stadtman TC, Rhee SG. Mammalian thioredoxin reductase: oxidation of the C-terminal cysteine/selenocysteine active site forms a thioselenide, and replacement of selenium with sulfur markedly reduces catalytic activity. Proc Natl Acad Sci U S A. 2000:97(6):2521–2526. https://doi.org/10.1073/pnas.050579797.
Loeb DD, Swanstrom R, Everitt L, Manchester M, Stamper SE, Hutchison CA III. Complete mutagenesis of the HIV-1 protease. Nature 1989:340(6232):397–400. https://doi.org/10.1038/340397a0.

Mariotti M, Ridge PG, Zhang Y, Lobanov AV, Pringle TH, Guigo R, Hatfield DL, Gladyshev VN. Composition and evolution of the vertebrate and mammalian selenoproteomes. PLoS One 2012:7(3): e33066. https://doi.org/10.1371/journal.pone.0033066.

Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, Tosatto SCE, Paladin L, Raj S, Richardson LJ, et al. Pfam: the protein families database in 2021. Nucleic Acids Res. 2021:49(D1):D412–D419. https://doi.org/10.1093/nar/gkaa913.

Molgedey L, Schuster HG. Separation of a mixture of independent signals using time delayed correlations. Phys Rev Lett. 1994:72(23): 3634–3637. https://doi.org/10.1103/PhysRevLett.72.3634.

Moshe A, Pupko T. Ancestral sequence reconstruction: accounting for structural information by averaging over replacement matrices. Bioinformatics. 2019:35(15):2562–2568. https://doi.org/10.1093/bioinformatics/bty1031.

Nguyen VD, Saaranen MJ, Karala A-R, Lappi A-K, Wang L, Raykhel IB, Alanen HI, Salo KEH, Wang C, Ruddock LW. Two endoplasmic reticulum PDI peroxidases increase the efficiency of the use of peroxide during disulfide bond formation. J Mol Biol. 2011:406(3): 503–515. https://doi.org/10.1016/j.jmb.2010.12.039.

Nyakatura K, Bininda-Emonds OR. Updating the evolutionary history of Carnivora (Mammalia): a new species-level supertree complete with divergence time estimates. BMC Biol. 2012:10(1):12. https://doi.org/10.1186/1741-7007-10-12.

Pollock DD, Thiltgen G, Goldstein RA. Amino acid coevolution induces an evolutionary stokes shift. Proc Natl Acad Sci U S A. 2012:109(21):E1352–E1359. https://doi.org/10.1073/pnas.1120084109.

Potter SC, Luciani A, Eddy SR, Park Y, Lopez R, Finn RD. HMMER web server: 2018 update. Nucleic Acids Res. 2018:46(W1): W200–W204. https://doi.org/10.1093/nar/gky448.

Rambaut A, Grassly NC. Seq-Gen: an application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. Comput Appl Biosci. 1997:13:235–238. https://doi.org/10.1093/bioinformatics/13.3.235.

Reich HJ, Hondal RJ. Why nature chose selenium. ACS Chem Biol. 2016:11(4):821–841. https://doi.org/10.1021/acschembio.6b00031.

Rennell D, Bouvier SE, Hardy LW, Poteete AR. Systematic mutation of bacteriophage T4 lysozyme. J Mol Biol. 1991:222(1):67–88. https://doi.org/10.1016/0022-2836(91)90738-R.

Romagné F, Santesmasses D, White L, Sarangi GK, Mariotti M, Hübler R, Weihmann A, Parra G, Gladyshev VN, Guigó R, et al. SelenoDB 2.0: annotation of selenoprotein genes in animals and their genetic diversity in humans. Nucleic Acids Res. 2014:42(D1): D437–D443. https://doi.org/10.1093/nar/gkt1045.

Sastry GM, Adzhigirey M, Day T, Annabhimoju R, Sherman W. Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. J Comput Aided Mol Des. 2013:27(3): 221–234. https://doi.org/10.1007/s10822-013-9644-8.

Scherer MK, Trendelkamp-Schroer B, Paul F, Pérez-Hernández G, Hoffmann M, Plattner N, Wehmeyer C, Prinz J-H, Noé F. PyEMMA 2: a software package for estimation, validation, and analysis of Markov models. J Chem Theory Comput. 2015:11(11): 5525–5542. https://doi.org/10.1021/acs.jctc.5b00743.

Sharir-Ivry A, Xia Y. Quantifying evolutionary importance of protein sites: a Tale of two measures. PLoS Genet. 2021:17(4): e1009476. https://doi.org/10.1371/journal.pgen.1009476.

Shema R, Kulicke R, Cowley GS, Stein R, Root DE, Heiman M. Synthetic lethal screening in the mammalian central nervous system identifies Gpx6 as a modulator of Huntington's disease. Proc Natl Acad Sci U S A. 2015:112(1):268–272. https://doi.org/10.1073/pnas.1417231112.

Steppan SJ, Adkins RM, Anderson J. Phylogeny and divergence-date estimates of rapid radiations in muroid rodents based on multiple nuclear genes. Syst Biol. 2004:53(4):533–553. https://doi.org/10.1080/10635150490468701.

Storz JF. Causes of molecular convergence and parallelism in protein evolution. Nat Rev Genet. 2016:17(4):239–250. https://doi.org/10.1038/nrg.2016.11.

Taylor A, Robson A, Houghton BC, Jepson CA, Ford WCL, Frayne J. Epididymal specific, selenium-independent GPX5 protects cells from oxidative stress-induced lipid peroxidation and DNA mutation. Hum Reprod. 2013:28(9):2332–2342. https://doi.org/10.1093/humrep/det237.

Tian R, Geng Y, Yang Y, Seim I, Yang G. Oxidative stress drives divergent evolution of the glutathione peroxidase (GPX) gene family in mammals. Integr Zool. 2021:16(5):696–711. https://doi.org/10.1111/1749-4877.12521.

Toppo S, Vanin S, Bosello V, Tosatto SCE. Evolutionary and structural insights into the multifaceted glutathione peroxidase (Gpx) superfamily. Antioxid Redox Signal. 2008:10(9):1501–1514. https://doi.org/10.1089/ars.2008.2057.

Tosatto SCE, Bosello V, Fogolari F, Mauri P, Roveri A, Toppo S, Flohé L, Ursini F, Maiorino M. The catalytic site of glutathione peroxidases. Antioxid Redox Signal. 2008:10(9):1515–1526. https://doi.org/10.1089/ars.2008.2055.

Trenz TS, Delaix CL, Turchetto-Zolet AC, Zamocky M, Lazzarotto F, Margis-Pinheiro M. Going forward and back: the complex evolutionary history of the GPx. Biology (Basel). 2021:10(11):1165. https://doi.org/10.3390/biology10111165.

Weinreich DM, Delaney NF, DePristo MA, Hartl DL. Darwinian evolution can follow only very few mutational paths to fitter proteins. Science 2006:312(5770):111–114. https://doi.org/10.1126/science.1123539.

Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 2007:24(8):1586–1591. https://doi.org/10.1093/molbev/msm088.

Yang Z, Wong WSW, Nielsen R. Bayes empirical Bayes inference of amino acid sites under positive selection. Mol Biol Evol. 2005:22(4):1107–1118. https://doi.org/10.1093/molbev/msi097.

Yates AD, Achuthan P, Akanni W, Allen J, Allen J, Alvarez-Jarreta J, Amode MR, Armean IM, Azov AG, Bennett R, et al. Ensembl 2020. Nucleic Acids Res. 2020:48(D1):D682–D688. https://doi.org/10.1093/nar/gkz966.

Zhang J, Kumar S. Detection of convergent and parallel evolution at the amino acid sequence level. Mol Biol Evol. 1997:14(5): 527–536. https://doi.org/10.1093/oxfordjournals.molbev.a025789.

Associate editor: Toni Gossmann