A peer-reviewed version of this preprint was published in PeerJ on 10 September 2015.

View the peer-reviewed version (peerj.com/articles/1244), which is the preferred citable publication unless you specifically need to cite this preprint.

https://doi.org/10.7717/peerj.1244
The selenium content of SEPP1 versus selenium requirements in vertebrates

Sam Penglase, Kristin Hamre, Ståle Ellingsen

Selenoprotein P (SEPP1) distributes selenium (Se) throughout the body via the circulatory system. The Se content of SEPP1 varies from 7 to 18 Se atoms depending on the species, but the reason for this variation remains unclear. Herein we provide evidence that vertebrate SEPP1 Sec content correlates positively with Se requirements ($R^2=0.88$). As the Se content of full length SEPP1 is genetically determined, this presents a unique case where a nutrient requirement can be predicted based on genomic sequence information.
The selenium content of SEPP1 versus selenium requirements in vertebrates.

S. Penglase¹,², K. Hamre¹ and S. Ellingsen¹

¹ National Institute of Nutrition and Seafood Research (NIFES), PO Box 2029, NO-5817 Bergen, Norway

² Department of Biology, University of Bergen, PO Box 7803, 5020 Bergen, Norway

Corresponding Author

PhD Sam Penglase

National Institute of Nutrition and Seafood Research (NIFES), PO Box 2029, 5817 Bergen, Norway

Telephone: +47 55905100

Fax: +47 55905299

E-mail: spe@nifes.no
Abstract

Selenoprotein P (SEPP1) distributes selenium (Se) throughout the body via the circulatory system. The Se content of SEPP1 varies from 7 to 18 Se atoms depending on the species, but the reason for this variation remains unclear. Herein we provide evidence that vertebrate SEPP1 Sec content correlates positively with Se requirements ($R^2=0.88$). As the Se content of full length SEPP1 is genetically determined, this presents a unique case where a nutrient requirement can be predicted based on genomic sequence information.
Selenium (Se) is an essential trace element required for the selenocysteine (Sec) residues inserted during mRNA translation into Se dependent proteins, termed selenoproteins (Brigelius-Flohé 1999). The number of selenoprotein coding genes differs among vertebrate species, ranging from the 24 to 25 found in mammals up to the 35 to 38 found in bony fish. Most selenoproteins are redox enzymes containing a single Sec residue involved in catalytic activity (Papp et al., 2007). An exception is the Se rich glycoprotein, selenoprotein P (SEPP1; aka SeP, SEPP, SEPP1a), which in vertebrates contains 7 to 18 Sec residues, depending on the species (Lobanov et al., 2008). The high Se content of SEPP1 is thought to facilitate Se distribution throughout the body.

In mammals, the liver is a major site of SEPP1 expression, where it is synthesised utilising food derived Se. Hepatic SEPP1 is then secreted into the blood plasma (Kato et al., 1992), where SEPP1 accounts for between 40 and 80% of the total Se (Hill et al., 1996; Hill et al., 2007; Read et al., 1990). Tissues utilise a combination of receptor mediated endocytosis and pinocytosis to obtain SEPP1 from the plasma, where it is then catabolised to release Se for de nova selenoprotein synthesis (Burk and Hill 2009; Burk et al., 2013).

Several features of SEPP1 are conserved among vertebrates including, i) a single N-terminal domain Sec residue present within a thioredoxin like motif (UXXC, where U is Sec), ii) a histidine rich region in the mid region of the protein), and iii) an apolipoprotein E receptor-2 (APOER2; aka LRP8) binding site followed by five Sec residues in proximity to the C-terminal (Figure 1) (Lobanov et al., 2008). APOER2 is widely expressed in human tissues (www.humanproteomemap.org; (Kim et al., 2014). APOER2 facilitated uptake of plasma SEPP1 is an essential (testes) or important (brain and foetus) pathway in some, but not all (muscle, kidney, liver or whole body) tissues for maintaining Se homeostasis in vivo (Burk et al., 2007; Burk et al., 2013; Hill et al., 2012; Olson et al., 2007). In contrast, the histidine rich regions of
SEPP1 presumably interact with multiple receptors, including megalin (LRP2). A megalin facilitated uptake pathway minimises excretion of Se by binding SEPP1 fragments in the kidney (Kurokawa et al., 2014; Olson et al., 2008) and plays a role in maintaining tissue Se homeostasis (Chiu-Ugalde et al., 2010; Steinbrenner et al., 2006). Additionally, the histidine rich regions are associated with the heparin binding properties of SEPP1. It is postulated that the heparin binding properties of SEPP1 allow the N-terminal Sec of SEPP1 to provide antioxidant protection for endothelial cells at sites of inflammation (Hondal et al., 2001; Saito et al., 2004).

In contrast, other domains in SEPP1 have low conservation among species. For instance, single base mutations in genomes have led to many cases of Sec to cysteine (Cys) substitution within the vertebrate SEPP1 C-terminal domain upstream and including the APOER2 binding site (Figure 1) (Lobanov et al., 2008). The reason why Sec content plasticity is observed only within this region of SEPP1 is unclear, but it is responsible for most of the variation between the SEPP1 Sec content among vertebrates (Lobanov et al., 2008). Furthermore, why SEPP1 Sec content differs among species also remains unknown. Several lines of evidence suggest vertebrate SEPP1 Sec number may be a direct function of Se utilisation. For instance, vertebrate SEPP1 Sec content correlates positively with selenoproteome size, tissue Se levels, and Se bioavailability in the environment (Lobanov et al., 2008).

If a direct relationship between SEPP1 Sec content and Se requirements exists, the SEPP1 Sec content of a species could predict its Se requirements, or vice versa. In doing so, this would provide a new insight into how the genome effects nutrient utilisation. Additionally, such a relationship would allow considerable scope for implementing the 3R’s (replace, reduce, refine). For example, this relationship would indicate the dietary Se levels to focus on when investigating the Se requirements for novel species. Such knowledge would reduce both the number of animals required and the risk of exposure to Se levels that may compromise animal welfare in such experiments.
In the following work, we compared the Sec content of mammalian and bony fish SEPP1’s predicted *in silico* with their Se requirements determined *in vivo*. We found a strong positive non-linear correlation (0.88) between the two, suggesting Se requirements can be predicted from the *Sepp1* gene sequence. The correlation was dictated by the Sec content within the C-terminal domain upstream and including the APOER2 binding site. The model was limited, whereby further analysis suggested it could not predict Se requirements in species whose SEPP1 Sec content was >16 residues, as found in many bony fish species. The predicted Se requirements for vertebrate species based on their SEPP1 Sec content are provided.
**Materials and methods**

The *in silico* predicted species specific Sec content of SEPP1 (SEPP1a in fish) were obtained from Lobanov et al. (2008), the open access selenoprotein database (selenodb.org; (Romagné et al., 2014)) or by analysing genomic *Sepp1* sequences (NCBI) for Sec content (http://seblastian.crg.es/), an open access online software for this purpose (Mariotti et al., 2013). The SEPP1 Sec content of five bony fish species; loach (*Paramisgurnus dabryanus*), cobia (*Rachycentron canadum*), grouper (*Epinephelus malabaricus*), gibel carp (*Carassius auratus gibelio*) and yellowtail kingfish (*Seriola lalandi*); were assumed to be within the 15 to 17 residue range found for fish in general (Lobanov et al., 2008) (See Supp. Table 2). The species specific Se requirement data were obtained from published studies and from the National Research Council of the USA (NRC) nutrient requirement reports (Gatlin and Wilson 1984; Han et al., 2011; Hao et al., 2014; Hilton et al., 1980; Jensen and Pallauf 2008; Le and Fotedar 2013; Lei et al., 1998; Lin and Shiau 2005; Liu et al., 2010; NRC 1963; 1985; 1995; 1997; 2011; Penglase et al., 2014; Sunde et al., 2009; Wedekind et al., 2004; Weiss et al., 1996; 1997). See Supp. Table 1 for further information regarding these animal Se requirement studies. Where multiple Se requirement studies for a species were available, the dietary Se requirements to fulfil the requirements of the actively growing juvenile stage was selected. Data were analysed in GraphPad Prism (GraphPad Software, San Diego, CA, USA, V. 5.04). Data were fitted with a horizontal line (null hypothesis) and then tested against more complex models in the following sequence; first order polynomial, second order polynomial and five parameter logistic equation (5PL) asymmetric sigmoidal; until the simplest model that explained the data was found (p<0.05). Model parameters were optimised to reflect current knowledge; vertebrates with seven Sec SEPP1; guinea pigs (*Cavia porcellus*) and naked mole rats (*Heterocephalus glaber*); have a Se requirement (Jensen and Pallauf 2008; Kasaikina et al., 2011). Thus the y intercept (no Se requirement) of models was constrained to ≤6
Sec for whole SEPP1 (Figure 2). Other vertebrate classes were excluded from the analyses because of limited data.

**Results and Discussion**

*The selenocysteine content of Selenoprotein P correlates strongly with selenium requirements.*

The Sec content in SEPP1 were identified for a total of 11 species; three bony fish and eight mammals; for which the Se requirements are also published (Supp. Table 1). Using this data, a positive non-linear correlation (R²=0.88) was found between the Se requirements and SEPP1 Sec number (Figure 2). Similarly, a positive correlation also occurs between SEPP1 Sec content and selenoprotein number in vertebrates (*Kryukov and Gladyshev 2000*; *Lobanov et al., 2008*). All fish annotated to date have SEPP1 (aka SEPP1a in fish) with 15 to 17 Sec residues (See Supp. Table 2). Based on this, an additional five bony fish species with known Se requirements were assumed to have SEPP1’s with 17 Sec residues and added to the data set, which was then re-analysed. This resulted in an asymmetric sigmoidal trend with a plateau at 17.0 (Figure 2), suggesting that a species SEPP1 is only useful for predicting Se requirements prior to this plateau (≤16 Sec residues). When a species SEPP1 has >16 Sec residues, as is found in many fish species, the curve predicts a minimum requirement (0.24 mg/Se kg dry matter (DM)) but not a maximum (there is no correlation between SEPP1 Sec content and Se requirements above this level). Modelling the data with alternative SEPP1 Sec content (15 or 16 Sec) for these five fish species shifts the plateau height towards those values, but retains the general features of the model (data not shown).

We then used this regression model (Figure 2) to predict the Se requirements of a species based on its SEPP1 Sec residue number predicted *in silico* (Table 1). As discussed before, there is a broad range of Se requirements found for bony fish that occurs in absence of an equally large distribution of SEPP1 Sec content. The reason for this occurrence is unknown. Perhaps the
relatively straightforward single base mutations of Sec to Cys codons (Lobanov et al., 2008) allowed mammalian SEPP1 Sec content to decrease in comparison to the ancestral vertebrate SEPP1 and in line with Se requirements, while fish utilised other regulatory mechanisms to increase Se supply to peripheral tissues. For example Sepp1 mRNA expression is elevated in fish, particularly in the kidneys, in comparison to mammals (Lobanov et al., 2008). This suggests plasma SEPP1 in fish may be replenished by SEPP1 synthesised from Se scavenged in the kidneys. As expected, the model (Figure 2) generally reflects the Se requirements of the species used to construct it, i.e. rat Se requirements are 0.1 mg Se/kg DM (Weiss et al., 1996; 1997) and it has a SEPP1 Sec content of 10 (Supp. Table 1), falling within the 0.08 ± 0.02 range predicted by the model (Table 1).

A hypothesis for the Sec number plasticity or conservation in different domains of vertebrate SEPP1.

As discussed, most of the difference in the SEPP1 Sec content between species is a result of differences in the Sec content found upstream and including the APOER2 binding site within the C-domain of SEPP1 (Figure 1 and Supp. Table 2). Thus as expected, when we analysed the Sec content in this region in relation to a species Se requirement (Supp. Figure 1), we found a similar positive correlation as found for full length SEPP1 and Se requirements (Figure 2). Recently it was found that SEPP1 Sec residues closer to the C-terminal are translated with greater efficiency than those towards the N-terminal (Shetty et al., 2014). Premature termination of SEPP1 translation at Sec codons appears to be a common event. For instance, four rat SEPP1 isoforms have been identified in plasma, whereby in addition to the full length protein, shorter variants are synthesised when translation is terminated at the second, third or seventh Sec codon (Ma et al., 2002). Thus, on average each plasma SEPP1 in mice contains 5 Sec residues, not the 10 Sec residues expected if only the full length protein is present (Hill et al., 2007). As a consequence of
this, a proportion of translated SEPP1 proteins will not contain the APOER2 binding site (Figure 1).

Taking in mind the above, we first hypothesise that decreases in Se requirements are an evolutionary adaption to Se availability. For instance, guinea pigs and naked mole rats both have low Se requirements (Jensen and Pallauf 2008; Kasaikina et al., 2011), and inhabit the Andes or East Africa respectively, both regions of low Se status (FAO 1992; Rachel et al., 2013). Secondly, we hypothesise that the Se requirements of the brain among species is similar on a weight basis, despite differences in the Se requirements of the whole body. For instance, compared to mice, naked mole rats have lower levels (-30 to -75%) of Se in most tissues except the brain (Kasaikina et al., 2011). And lastly, low Se availability can stall translation of selenoproteins at Sec codons (Weiss Sachdev and Sunde 2001), and perhaps results in the truncated forms of SEPP1 translated in vivo. We therefore hypothesise that Sec to Cys substitutions in SEPP1 occurred specifically in the region downstream and including the APOER2 binding site as it aids the translation of full length protein under Se limiting conditions, such as those faced by naked mole rats and guinea pigs. The subsequent retention of the APOER2 binding site would allow the continuation of a controlled Se supply to critical organs utilising APOER2 mediated uptake of SEPP1, such as the brain.

Conclusion

The Sec content of SEPP1 correlates with Se requirements in vertebrates when Sec residue number is \( \leq 16 \). There was no correlation between SEPP1 Sec content and Se requirements for species with SEPP1’s with >16 Sec residues, as is the case for many bony fish species. However, for those species with SEPP1’s with >16 Sec residues, a minimum Se requirement of 0.24 mg Se/kg DM was predicted. This study suggests that genome evolution is affected directly by
nutrient availability in the environment, and provides novel evidence that the genomic sequence can be used to predict a nutrient requirement.

References


Steinbrenner H, Bilgic E, Alili L, Sies H, Brenneisen P. 2006. Selenoprotein P protects endothelial cells from oxidative damage by stimulation of glutathione peroxidase expression and activity, **40**:936-943. doi:10.1080/10715760600806248.

Sunde RA, Raines AM, Barnes KM, Evenson JK. 2009. Selenium status highly regulates selenoprotein mRNA levels for only a subset of the selenoproteins


Table 1 (on next page)

The Se requirements (mg Se/kg DM) predicted by the model (Figure 2) versus those determined in feeding trials for vertebrates as the selenocysteine content (Sec no.) of selenoprotein P (SEPP1) increases.

Predicted Se requirement data are mean ± 95% confidence interval. \(^1\) mg Se/kg feed DM, mean (± 95% confidence interval, when shown) \(^2\) Further information on the representative species used for the determined Se requirement data can be found in Supp. Table 1. \(^3\) There are currently no known species with full length SEPP1 containing 6 Sec residues.
**Table 1.** Table of the Se requirements (mg Se/kg DM) predicted by the model (Figure 2) versus those determined in feeding trials for vertebrates as the selenocysteine content (Sec no.) of selenoprotein P (SEPP1) increases. Predicted Se requirement data are mean ± 95% confidence interval.

<table>
<thead>
<tr>
<th>Class</th>
<th>Sec no.</th>
<th>Predicted Se requirement&lt;sup&gt;1&lt;/sup&gt;</th>
<th>Determined Se requirement&lt;sup&gt;1&lt;/sup&gt;</th>
<th>Representative species&lt;sup&gt;2&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bony</td>
<td>6</td>
<td>0.02 ± 0.03</td>
<td>ND</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>0.04 ± 0.03</td>
<td>0.06</td>
<td>Guinea pig</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>0.05 ± 0.02</td>
<td>ND</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>0.07 ± 0.02</td>
<td>ND</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>0.08 ± 0.02</td>
<td>0.10</td>
<td>Rat</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>0.10 ± 0.02</td>
<td>ND</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>0.12 ± 0.03</td>
<td>0.10</td>
<td>Cow</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>0.14 ± 0.04</td>
<td>0.10</td>
<td>Horse</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>0.17 ± 0.05</td>
<td>0.20</td>
<td>Pig</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>0.20 ± 0.06</td>
<td>0.21</td>
<td>Dog</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>0.24 ± 0.06</td>
<td>0.25</td>
<td>Catfish</td>
</tr>
<tr>
<td></td>
<td>17+</td>
<td>&gt;0.24</td>
<td>0.30</td>
<td>Zebrafish</td>
</tr>
</tbody>
</table>

<sup>1</sup> mg Se/kg feed DM, mean (± 95% confidence interval, when shown)

<sup>2</sup> Further information on the representative species used for the determined Se requirement data can be found in Supp. Table 1.

<sup>3</sup> There are currently no known species with full length SEPP1 containing 6 Sec residues.
The receptor binding sites and selenocysteine (Sec) residues of vertebrate selenoprotein P (SEPP1)

From the N-terminal side, SEPP1 is comprised of a conserved N-terminal domain Sec residue, followed by several proposed heparin binding sites which include a histidine rich region. Following this, there is the shorter Sec residue rich C-terminal domain which contains an APOER2 binding site. The C-terminal domain can be further divided into two subdomains. The first subdomain exists on the N-terminal side of the APOER2 binding site and contains a region with a low conservation of Sec residues among vertebrates (mainly due to Sec to cysteine (Cys) conversions (Lobanov et al., 2008)). The second subdomain is located downstream of the APOER2 binding site and contains five Sec residues that are conserved across vertebrate species. Several species of amphibians also have an additional Sec residue in the C-terminal end of this region (Lobanov et al., 2008). The proposed heparin binding sites/histidine rich regions are based on rat SEPP1 found by Hondal et al. (2001). Similar histidine rich regions are found in the SEPP1’s of other species (selenodb.org). Cys residues outside the C-terminal domain are not shown. Red lines = conserved Sec residues; Black lines = Cys or Sec residues; Green lines = Cys/Sec residues within the APOER2 binding site; Green box grids = proposed heparin binding sites.
2

The relationship between the selenocysteine content of selenoprotein P and selenium requirements in vertebrates.

The solid line with the solid circles (●) is the best fit model for the SEPP1 Sec content versus Se requirements (mg Se/kg dry matter (DM)) from 11 species with representatives from the mammalian and bony fish classes where the genome sequences were available (Second order polynomial, $R^2 = 0.88$, $y = 4.3 + 78x - 122x^2$). The broken line represents the same data modelled with an additional five bony fish species with known Se requirement levels (○), but unannotated genomes. SEPP1 Sec content in these fish were assumed to be within the likely range of 15-17 Sec residues found for fish in general (5PL Asymmetric sigmoidal, $R^2 = 0.92$, $y = 5.13 + (11.9/((1+10^{((X-1.6410)\times 6.391)})9.611^{10}))$. Shaded boxes group animals within classes. The X axis is log transformed.