Hepatitis C virus encodes a selenium-dependent glutathione peroxidase gene. Implications for oxidative stress as a risk factor in progression to hepatocellular carcinoma

By: Weiqing Zhang, Arthur G. Cox & Ethan Will Taylor

Weiqing Zhang, Arthur G. Cox & Ethan Will Taylor. Hepatitis C virus encodes a selenium-dependent glutathione peroxidase gene. Implications for oxidative stress as a risk factor in progression to hepatocellular carcinoma *Medizinische Klinik* **94,** 2 (1999). https://doi.org/10.1007/BF03042181

Abstract:

Aim: Using structural bioinformatics methods, the aim is to assess the hypothesis that hepatitis C virus (HCV) encodes a glutathione peroxidase (GPx) gene in an overlapping reading frame, linking HCV expression and pathogenesis to the Se status and dietary oxidant/antioxidant balance of the host. *Methods:* The putative HCV GPx gene was identified by searching viral sequence databases, using conserved GPx active site sequences as probes, giving particular weight to the UGA (selenocysteine) codon. Multiple sequence alignments were generated and analyzed to validate the sequence similarity, and to establish the degree of conservation of the identified genomic features in HCV. Molecular modeling was used to assess the structural feasibility of the proposed homology. *Results:* The GPx homology region overlaps the NS4 gene, and is well conserved in HCV. The sequence similarity of the conserved active site regions to a set of known GPx is high (4 to 6 SD greater than expected for similar random sequences). The computed strain energy of a molecular model of the HCV GPx is energetically favorable, comparable to the bovine GPx structure. *Conclusions:* By linking HCV replication and pathogenesis to the Se status and dietary oxidant/antioxidant balance of the host, the existence of a viral GPx gene could help to explain why HCV disease progression is accelerated by oxidant stresses such as alcoholism and iron overload.

Keywords: Gluthathione Peroxidase | Hepatitis C virus | Hepatecullular carcinoma | Selenium | Selenoprotein

Article:

The possibility that viruses might encode selenoproteins remained unexplored until 1994 when Taylor et al. [10] published a study of the predicted RNA structure of the human immunodeficiency virus, HIV-1, in relation to potential novel open reading frames (ORFs, protein coding regions) of the virus. That analysis demonstrated the potential for several new

gene variants in HIV-1, which might encode selenoproteins, because of the existence of several highly conserved UGA codons (which can encode selenocysteine) in certain regions of HIV. These UGA codons were consistently found to be downstream of potential RNA structural features ("slippery" sequences and RNA pseudoknots) that would be required for the expression of these hypothetical genes by ribosomal frameshifting.

Subsequently, the hypothetical protein encoded by one of these predicted HIV-1 selenoprotein genes was shown to have highly significant sequence similarity to the active site domains of the mammalian selenoprotein glutathione peroxidase (GPx), despite being truncated by several internal and C-terminal deletions [8]. Along the same lines, highly truncated GPx-related sequences were identified in the potentially cardiovirulent Se-dependent B3 strain of coxsackievirus, CVB3 [9], which is the same strain that has been studied by Beck et at. [1] as a putative Keshan disease cofactor.

The only experimentally verified example of a viral selenoprotein is from a DNA virus, the pox virus *Molluscum contagiosum,* which has been shown to encode a functional full length GPx enzyme [7]. Zhang et al. [11] subsequently reported that a number of RNA viruses encode GPx-related sequences, including, in addition to the HIV-1 and CVB3 examples mentioned above, potential GPx genes in measles virus and the hepatitis C virus, HCV. Thus, it seems likely that this antioxidant selenoprotein module may prove to be a relatively common gene in both RNA and DNA viruses.

The current paper focuses on the case of the hypothetical HCV GPx gene first predicted by Zhang et al. [11], which is of considerable interest because HCV infection is very prevalent worldwide, is considered an emerging public health threat, and has poorly understood mechanisms of pathogenesis. We will assess the hypothesis that the HCV encoded sequence is a GPx homologue by means of structural bioinformatics methods, including some preliminary results of homology molecular modeling studies of the HCV GPx protein structure, which will be reported in depth in a separate publication [3]. Finally, we will review clinical data and other lines of evidence that are consistent with this hypothesis, and consider its clinical implications.

METHODS

Database Searches and Potential Gene Identification

As described previously by Zhang et al. [11], the potential HCV glutathione peroxidase (GPx) gene was identified by searching sequence databases with the program TFASTA, using known variants of several different GPx active site sequences as probes, e. g., VASYUGLT (plasma GPx), etc., where U represents the TGA codon in DNA, potentially encoding selenocysteine (Sec). An altered translation table was used, modified specifically so as to distinguish the TGA codon from the other 2 stop codons, as described in detail previously [11]. Highly ranked database hits were subsequently examined by translating all 3 reading frames in the region of the GPx sequence similarity, to determine the potential for expression of the potential gene by known mechanisms, i.e. a start codon, ribosomal frameshift site, or an RNA editing site.

Sequence Alignment and Assessment of Statistical Significance of Sequence Similarities

The BlockAlign program [11] was used to generate multiple sequence alignments; this program produces an optimal, gapped alignment of a probe sequence against an existing sequence alignment. Combined with a randomization routine, it was also used for assessing the statistical significance of the similarity between a query sequence and a prealigned protein family (in this case, a multiple alignment of GPx sequences). Briefly, the method involves 1. the optimal alignment of the query sequence against the prealigned family (GPx), which also yields a similarity score, followed by 2. repeated random shuttling and realignment of the query sequence, 3. calculation of the average similarity score and standard deviation for optimally aligned random sequences of identical composition to the query sequence. The significance of the query-family homology can then be expressed as a shuffling statistic (Z score), calculated as the distance of the actual score from the average random score, in standard deviations (SD). The blosum62 amino acid similarity matrix was used for all sequence alignments, and Sec residues were treated as Cys for the purpose of these calculations.

Molecular Modeling

The Sybyl program (Tripos Assoc., St. Louis, MO) was used to build a putative HCV viral GPx homology model, starting from the bovine cellular GPx X-ray crystal structure, file lGPl from the Brookhaven Protein Data Base. The model was constructed based on the alignment shown in simplified form in Figure 1 of Zhang et al. [11]. The model was constructed using the biopolymer protein-loop option of Sybyl, which uses a knowledge-based approach to rebuild regions where insertions and deletions are predicted by the alignment.

The resulting model was refined by energy minimization using molecular mechanics; the final minimization used the Kollman all-atom force field as implemented in Sybyl. The Sec residue was modeled as the deprotonated Se-species. Default program options were used for the minimization, with a ΔE of 0.001 kcal/mol as the termination criterion in the final energy minimization. A monomer from the bovine GPx structure was minimized similarly for comparison. Additional details regarding these computational studies, the parameters used, as well as further optimization and analysis of the HCV viral GPx model using molecular dynamics, are reported in a separate publication [3].

RESULTS AND DISCUSSION

Virus sequence database searches using variants of the Sec encoding region of GPx led to the identification of the GPx homology region in HCV, where the hypothetical GPx gene is in the -1 reading frame overlapping a known nonstructural gene, NS4b, of previously unidentified function. The overlapping region contains one in-frame "stop" codon, UGA, that can potentially encode Sec; this region also lacks a start codon. This may help to explain why this putative gene has not been noted previously, because it does not look like an open reading flame, and a known gene is encoded in the same region of RNA. The Se-dependent GPx sequence and UGA codon are well conserved in HCV genotype 1b (which is predominant in North America), and also in subtypes 2b and 3a. In several other HCV subtypes (la, lc), the UGA codon has mutated to an AGA (arginine) codon, a mutation which may be consistent with a non-Se-dependent

GPx activity. The same mutation was noted in the putative HIV-1 GPx gene, in subtypes prevalent in Africa and Asia, whereas the UGA codon (Se-GPx) was conserved in the sub)-pes predominant in Europe and the Americas [8].

As shown in the simplified multiple sequence alignment of Figure 1, the putative HCV GPx sequence has regions of substantial similarity to known GPx sequences; the similarity encompasses the entire sequence, but is strongest in the active site regions. For a sequence block of about 25 residues (beginning VNVAS...) spanning active site region l, the similarity is statistically significant at > 6 SD relative to random sequences of similar composition. The HCV GPx has features of both the plasma and cellular GPx types.

```
          #                                  #                                    #
    V*VAS UGL G* *VEL*  QR*  PR* *I G*  * * HQ PA*  E  P * **R* GG  P WE* CRH   *WS   *   G
V1  VQVAGPUGLLGKAHVELH..QRDT@PRDSITDGI..HSLYHQ.SAHHPEY.PPVQHLRGMGG.CS%WEGACGHSGGLWSGGGRCTRGF
V2  IQVASPUGFLGEAHVEFH..QRDT@PRDSIIDGI..HSLHHQ.PAHHPKH.PPVQHLGRVGG.RP%WEGACRHSGGLWSRSGGCSRGL
P1  VNVASYUGLTD.QYLELNALQEELGPFGLVILGFPSNQFGKQEPGENSEILPSLKYVRPGGGFVP%WEPMKIHDIR.WNFE.KFLVGP
P2  VNVASYUGLTD.QYLELNALQEELGPFGLVILGFPCNQFGKQEPGENSEILPSLKYVRPGGGFVP%WEPMKIHDIR.WNFE.KFLVGP
P3  VNVASYUGLTG.QYIELNALQEELAPFGLVILGFPCNQFGKQEPGENSEILPTLKYVRPGGGFVP%WEPMKVHDIR.WNFE.KFLVGP
P4  VNVASYUGLTG.QYVELNALQEELEPFGLVILGFPCNQFGKQEPGENSEILATLKYVRPGGGFTP%WEPMKVHDIR.WNFE.KFLVGP
C1  ENVASLUGTTIRDYTEMNDLQKRLGPRGLVVLGFPCNQFGHQENGKNEEILNSLKYVRPGGGFEP%WSPVCRNDIA.WNFE.KFLVGP
C2  ENVASLUGTTVRDYTQMNDLQRRLGPRGLVVLGFPCNQFGHQENAKNEEILNCLKYVRPGGGFEP%WSPVCRNDVS.WNFE.KFLVGP
C3  ENVASLUGTTVRDYTQMNELQRRLGPRGLVVLGFPCNQFGHQENAKNEEILNSLKYVRPGGGFEP%WSPVCRNDVA.WNFE.KFLVGP
C4  ENVASLUGTTVRDYTQMNELQERLGPRALVVLGFPCNQFGHQENAKNEEILNSLKYVRPGGGFEP%WCPVCRNDVS.WSFE.KFLVGP
    └────R1─────┘                        └────R2─────┘                               └─R3─┘
```

Figure 1. Simplified multiple sequence alignment (single letter amino acid code) of the HCV encoded GPx homologue vs. a set of Se-GPx sequences. Two HCV genotype 1b strains are shown as V1 and V2; P1–P4 are plasma GPx, C1–C4 are cellular GPx. The symbol # indicates the positions of the conserved catalytic triad of Sec (U), Gln (Q) and Trp (W) in active regions R1 to R3. Letters on the line above the alignment indicate an exact match between an HCV residue and one or more of the mammalian GPx sequences; an asterisk indicates a chemically similar residue. For brevity, 2 regions of the alignment are not shown: an N-terminal region of 35-40 residues and an internal region at %, of 25-40 residues. The symbol @ indicates an insertion of 11 amino acids in HCV, containing a conserved Trp residue, which, in many HCV isolates where this reading frame is more severely truncated at the 3′ (C-terminal) end, may serve to substitute for the function of the active site Trp in R3.

As discussed in detail previously [11], the HCV GPx is probably expressed via 1LNA editing rather than the ribosomal frameshift mechanism proposed for the HIV-1 GPx gene [10]. The putative RNA editing site in HCV, consisting of a homopyrimidine sequence (CCUCCCUU) with a triplet of C bases, is just downstream of a known HCV protease cleavage site (FDEMEEC/ASHLP), at the NS4a-NS4b junction in the main polyprotein open reading frame. This implies that, when translated front an edited RNA, the HCV GPx would be cleaved from the viral polyprotein only a few residues upstream from the beginning of the GPx homology region. This processing to form an independent protein suggests a predominantly intracellular and possible intravirion role for the HCV GPx.

A molecular model of the putative HCV GPx homologue was constructed using the HCV subtype lb sequence shown in Figure 1 as sequence V2 (Genbank accession #D504830. There is one insertion of about 10 residues in HCV (at the position shown by the symbol @ in Figure 1), which contains several highly conserved residues, including an Arg and a Trp, that are likely to

play an important role in the viral GPx homologue. However, because this has no direct structural equivalent in the bovine GPx structure, we did not attempt to model this insertion explicitly. All of the other insertions and deletions proved to be structurally feasible, because of extensive coiling of the protein backbone in the affected regions. Minimization of the protein model lead to a substantially negative molecular mechanics "strain" energy of -2,002 kcal/mole for 133 residues (-15.1 kcal/mole·residue), which on a per-residue basis is reasonably close to, but not quite as low as, the value obtained by minimizing the bovine GPx crystal structure, -3,619 kcal/mole for 185 residues (-19.6 kcal/mole·residue). Part of the difference arises because the HCV GPx has less electrostatic stabilization than the bovine GPx because the former has a net positive charge, which might be functionally significant (e.g., suggesting association of the viral GPx with the negatively charged backbone of the viral RNA in the virion, or with phospholipid membrane of the viral envelope or infected cell). Furthermore, the geometry of the GPx active site, and distances between active site residues, are not substantially changed in the model (Figure 2). Thus, overall, the molecular modeling results suggest that the proposed homology is structurally feasible.

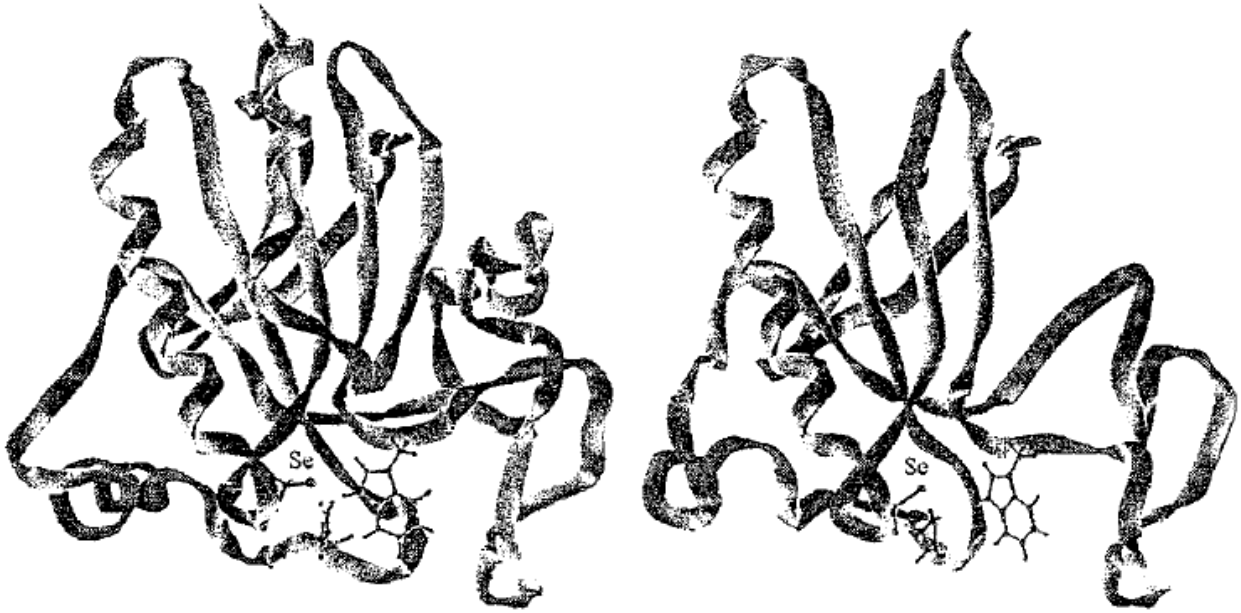

Figure 2. A comparison of the protein backbone and active site residues of the bovine GPx monomer (left) and a homology molecular model of the putative HCV viral GPx enzyme (right). The protein backbone is shown as a ribbon, and the sidechains of the catalytic triad of Sec, Gln and Trp are shown using a ball and stick rendition. The Se atom of the Sec residue is indicated.

Several additional lines of evidence support the hypothesis that Se status may play a role in HCV disease progression and outcome:

1. The best direct evidence consistent with an HCV-Se link are the clinical data of Look et al. [4], who found that in HIV+ patients, the progressive decline in Se levels characteristic of HIV infection was greater in those with HCV co-infection, who "showed markedly lower Se concentrations compared to those without concomitant HCV-infection" [4].

2. HCV infection has been associated with dilated cardiomyopathy consequent to viral myocarditis [9]; this is similar to Keshan disease, the classical Se-deficiency disease, which is characterized by cardiomyopathy. Keshan disease is believed to involve a viral cofactor, coxsackie virus, which shows increased cardiovirulence in Se deficient hosts [1]. Cardiomyopathy has also been linked to Se deficiency in HIV infection [10].
3. Oxidant stresses like alcoholism and iron overload are associated with HCV disease progression, suggesting an antioxidant impairment is involved in pathogenesis. This iron effect is so significant that phlebotomy or intentional bleeding has been found to significantly improve recovery and response to interferon treatment in HCV infection [6 I. Unbound iron is a powerful generator of reactive oxygen species via the Fenton reaction. Se protects against such free radicals and peroxides, and Se sequestration in a viral protein could make the host more susceptible to oxidative stress.
4. In several studies, people of low socioeconomic status were observed to have the highest risk of HCV infection; this is consistent with the possibility that malnutrition and consequent Se deficiency can enhance vulnerability to HCV infection, or increase chronicle of the infection by making it more difficult to overcome a primary infection.

In regard to the question of what a vires might gain by encoding a GPx, there are a number of possibilities: 1. A vitally-encoded GPx could help a virus defend against free radical mediated attacks on infected cells by the immune system. 2. If associated with the viral particle, a viral GPx could also increase the extracellular viability of virus particles in the blood and extracellular compartments, at least in the case of an enveloped virus such as HCV, which is more susceptible to membrane lipid peroxidation once it has budded off the host cell and has lost the benefit of cellular antioxidant defenses. 3. A viral GPx might also have a regulatory (and potentially repressive) effect on viral replication, because, for example, it is known that oxidative stress (e.g. H_2O_2 exposure) activates the replication of HIV-1 and other viruses: a viral GPx would reduce oxidant tone, thus reducing viral activation. 4. A viral GPx might play a role in attachment to or penetration of target cells.

Whatever the specific benefits that the virus may gain by encoding a Se-dependent GPx gene, it is likely that one mechanism of viral pathogenesis resulting from the expression of that gene would be competition between viral and host selenoprotein synthesis, particularly under conditions where dietary Se intake is suboptimal, as is the case in many human populations. Stated differently, the expression of this viral gene could lead to a decline in cellular selenoprotein levels, and thus make the host more susceptible to the effects of Se deficiency. !L as suggested above, GPx can act as a negative regulator of viral replication (which has been established in the case of HlV-1 [11]), then Se deficiency might also be associated with enhanced viral replication, and higher viral loads. In contrast, supplementation with Se should act to minimize the negative impact of viral GPx expression on host selenoprotein expression, while slowing viral replication. In essence, by this hypothesis, the effect of Se on HCV infection would be to tend to stablish a homeostasis between the virus and the host. Other antioxidants that are either Se-sparing (e.g. vitamin E) or that boost glutathione levels (e.g. lipoic acid) could act similarly to or synergistically with Se in HCV infection. It must be emphasized that these hypotheses as to potential antiviral mechanisms of Se and other antioxidants vs. HCV are purely speculative and have yet to be definitively tested by in-vitro or in-vivo studies. However, the

question of mechanism aside, there do appear to be clinical benefits associated with antioxidant therapy of HCV infection [2].

CONCLUSIONS

The prediction of GPx coding potential in HCV is primarily based upon the presence and conservation of significant protein sequence homology, particularly in active site regions, combined with evidence that the gene region in question may be expressed by known mechanisms of RNA editing and readthrough suppression. The specific viral mechanisms or RNA structure features that might be involved in recoding the UGA codon for translation as Sec remain to be identified. The results of the molecular modeling studies, and a number of independent lines of clinical evidence are also consistent with the hypothesis.

The existence of an HCV encoded Se-GPx, well conserved in genotype 1b, may help to explain certain clinical aspects of HCV infection. Significantly, genotype 1b is associated with the highest risk of progression to cirrhosis and hepatocellular carcinoma, and poor response to interferon. The outcome of chronic HCV infection is currently difficult to predict, and many carriers never manifest symptoms of disease, which begs the question as to what individual or environmental factors determine that outcome Our results suggest that correlations between HCV disease progression and factors such as dietary Se intake and blood levels, markers of lipid peroxidation, etc., should be systematically investigated as potential correlates and predictors of outcome m HCV infection.

In the light of our findings, it is of considerable interest that Berkson [2] has reported the successful treatment of individual cases of HCV-associated viral hepatitis with an antioxidant regimen that includes lipoic acid and Se. Thus, the potential of dietary antioxidants as primary or complementary therapies for HCV infection merits serious investigation.

REFERENCES

1. Beck MA, Siu Q, Morris VC, et al. Rapid genomic evolution off a non-virulent coxsackievirus B3 in selenium-deficient mice results in selection of identical virulent isolates Nature Med 1995; 1:433-6.
2. Berkson BM. A conservative triple antioxidant approach to the treatment of hepatitis C that combines aopha lipoic acid (thioctic acid), silimarin, and selenium. Three case histories. Med Kiln 1999;94:Suppl III:84-9.
3. Cox AG, Zhang W, Taylor EW. Molecular modeling and comparative sequence analysis support the hypothesis that hepatitis C virus encodes a selenium-dependent glutathione peroxidase gene. (In preparation.)
4. Look MP, Rockstroh JK, Rao GS, et al. Serum selenium, plasma glutathione (GSH) and erythrocyte glutathione peroxidase (GSH-Px)-levels in asymptomatic versus symptomatic human immunodeficiency virus-1 (HIV-l)-infection. Eur J Clin Nutr 1997:51:266-72.
5. Matsumori A, Matoba Y, Sasayama S. Dilated cardiomyopathy associated with hepatitis C virus infection. Circulation 1995;92:2519-25.

6. Shedlofsky SI. Role of iron in the natural history and clinical course of hepatitis C disease. Hepatogastroenterology 1998;45:349-55.

7. Shisler JL, Senkevich TG, Berry MJ, et al. Ultraviolet-reduced cell death blacked by a selenoprotein from a human dermatotropic poxvirus. Science 1998;279:102-5.

8. Taylor EW, Bhat AA, Nadimpalli RG, et al, HIV-1 encodes a sequence overlapping env gp41 with highly significant similarity to selenium-dependent glutathione peroxidases. J Acquir Immune Defic Syndr Hum Retrovirol 1997; 15:393-4

9. Taylor EW, Nadimpalli RG, Ramanathan CS. Genomic structures of viral agents in relation to the biosynthesis of selenoproteins. Biol Trace Elem Res 1997;56:63-91.

10. Taylor EW, Ramanathan CS, Jalluri RK. et al. A basis for new approaches to the chemotherapy of AIDS: novel genes in HIV-1 potentially encode selenoproteins expressed by ribosomal frameshifting and termination suppression J Med Chem 1994;37:2637-54.

11. Zhang W, Ramanathan CS, Nadimpalli RG, et al. Selenium-dependent glutathione peroxidase modules encoded by RNA viruses. Biol Trace Elem Res I999:72:97-116.