Selenium-dependent glutathione peroxidase modules encoded by RNA viruses

By: W. Zhang, C. S. Ramanathan, R. G. Nadimpalli, A. A. Bhat, A. G. Cox & E. W. Taylor

W. Zhang, C. S. Ramanathan, R. G. Nadimpalli, A. A. Bhat, A. G. Cox & E. W. Taylor. Selenium-dependent glutathione peroxidase modules encoded by RNA viruses. *Biological Trace Element Research* **70,** 97–116 (1999). https://doi.org/10.1007/BF02783852

Abstract:

Glutathione peroxidase (GPx) is the prototypical eukaryotic selenoprotein, with the rare amino acid selenocysteine (Sec) at the enzyme active site, encoded by the UGA codon in RNA. A DNA virus, *Molluscum contagiosum*, has now been shown to encode a functional selenium-dependent GPx enzyme. Using modifications of conventional sequence database searching techniques to locate potential viral GPx modules, combined with structurally guided comparative sequence analysis, we provide compelling evidence that Se-dependent GPx modules are encoded in a number of RNA viruses, including potentially serious human pathogens like HIV-1 and hepatitis C virus, coxsackievirus B3, HIV-2, and measles virus. Analysis of the sequences of multiple viral isolates reveals conservation of the putative GPx-related features, at least within viral subtypes or genotypes, supporting the hypothesis that these are functional GPx modules.

Keywords: AIDS | genomic analysis | glutathione peroxidase | hepatitis C virus | HIV | measles virus | retrovirus | selenium | selenoproteins

Article:

INTRODUCTION

Selenium (Se) is an essential trace mineral, which can be specifically incorporated into selenoproteins as the rare amino acid selenocysteine (Sec, the Se analog of cysteine); Se can also be incorporated nonspecifically as selenomethionine (SeMet) into "Se-containing" proteins, which must be distinguished from true selenoproteins *(1).*

In mammals, Se is known to be critical for antioxidant defenses, thyroid and reproductive functions, and aspects of immune function, particularly cellular immunity *(1-3).* Most of these important functions of Se are associated with specific selenoproteins. Se is critical for antioxidant defenses largely because it is an essential component of many glutathione peroxidases (GPx), although non-Se-dependent forms of GPx also exist, most commonly with

the active site Sec replaced by Cys *(1).* By reducing cytosolic and extracellular peroxides, the various forms of this enzyme are essential for combating the ubiquitous and generally harmful process of lipid peroxidation, which is increased under conditions of "oxidative stress." Unreversed lipid peroxidation leads to cell-membrane destruction and can be induced during apoptosis (programmed cell death), showing that free-radical activity is not always harmful to organisms if it can be properly compartmentalized *(4).*

Given the evidence for the important role of Se in immune function *(1,5),* related in part to redox regulation of cellular processes like apoptosis used by immune cells *(4),* it is clear that a virus might benefit from being directly involved in these processes by encoding a selenoprotein *(3,6).* However, it would have to encode this as a highly regulated auxiliary gene usually expressed at low levels, in order not to severely deplete cellular Se and prematurely induce free-radical-mediated cell death, which would be counter to the interests of the virus, which depends on the survival of the host cell for its replication.

Genomes of representative bacteria, archea, and eukaryotic organisms are known to encode selenoproteins; all use the UGA codon (normally a protein termination signal) for *Sec,* and a complex cotranslational mechanism for Sec incorporation into polypeptide chains, involving RNA stem-loop structures *(7-9).* However, the possibility that viruses might encode selenoproteins remained unexplored until Taylor et al. published a study of the predicted RNA structure of HIV in relation to potential novel open-reading frames (ORFs, protein-coding regions) of the virus *(10).* This analysis demonstrated the potential for several new gene variants in HIV that possibly encode selenoproteins, because of the existence of several highly conserved UGA codons in certain regions of HIV, downstream of potential RNA structural features ("slippery" sequences and RNA pseudoknots) that would be required for the expression of these genes by ribosomal frameshifting. In subsequent studies *(6,11,12),* these theoretical findings were extended, and the same approach was applied to other viruses, including the picornavirus coxsackie B3 (CVB3), the putative cofactor in Keshan disease.

Keshan disease is a classical Se-deficiency disease, named after a county in China where outbreaks occurred because of the very low Se levels in soils of the region. Women and young children seem particularly susceptible to the disease, of which the primary symptom is a nonobstructive cardiomyopathy, leading to death by heart failure in severe cases. Because of the seasonal and clustered nature of outbreaks of the disease, Chinese investigators suspected the involvement of an infectious agent or other cofactor and eventually isolated coxsackievirus from the hearts of disease victims. The role of coxsackievirus in Keshan disease is strongly supported by demonstrations that a deficiency of Se can trigger a similar cardiomyopathy in coxsackie-infected mice *(13).* Beck and co-workers have confirmed this finding and have shown that even a "benign" strain of coxsackievirus B3 (CVB3) becomes virulent in Se-deficient animals *(14),* where, undergoing specific genomic changes, it mutates into a more virulent strain that can produce myocarditis even in Se-adequate mice *(15).*

The possibility of virally encoded selenoproteins has now been firmly substantiated by the findings of Moss and co-workers, who in 1996 reported the complete genomic sequence of the pox virus *Molluscum contagiosum (16),* in which they identified an open-reading frame (ORF) highly homologous to the known mammalian selenoprotein GPx, with 76% sequence identity at

the amino acid level, with an identically placed in-frame UGA codon. This gene was later shown to encode a functional GPx enzyme *(17).*

Taylor et al. also identified highly distinctive but substantially truncated GPx-related sequences in the potentially cardiovirulent Se-dependent B3 strain of coxsackie virus *(6),* which is the same strain that has been studied by Beck and co-workers as a putative Keshan disease cofactor *(14,15).* Subsequently, one of the potential selenoprotein genes predicted previously in HIV-1 by Taylor et al. *(10)* was also shown to encode a GPx homolog, which although truncated by several substantial deletions nonetheless includes the essential active site regions and catalytic amino acids of the GPx enzyme *(18).*

These preliminary findings led us to undertake a systematic search for GPx modules in viral genomes, expecting that this antioxidant selenoprotein module may ultimately prove to be a relatively common component of a number of RNA and DNA viruses, which, indeed, seems to be the case, as our present results will demonstrate. The current article focuses on examples of potential GPx genes or modules that we have identified in a number of clinically significant RNA viruses, including a review and update on two previously published cases, in HIV-1 and CVB3, as well as the first report of an unmistakable GPx homolog encoded in an overlapping reading frame of hepatitis C virus and several other potential GPx modules with highly conserved UGA codons in GPx-like motifs, in measles virus, and at the 3' end of the HIV-2 envelope gene.

METHODS

There are certain difficulties or limitations in the use of standard database search tools for the identification of potential selenoprotein genes in existing genomic sequence data. First, amino acid mutation matrices commonly used for database searching and sequence alignment are not parameterized for Sec, so the use of standard search tools requires treatment of Sec in the probe sequence as a Cys residue, which eliminates much of the specificity for selenoproteins. Furthermore, in protein databases, selenoprotein sequences are often entered using either an X or a C (Cys) for any Sec residues, so searches can potentially miss or underscore a hit if an X is used in a database entry. Finally, if a search that does a translation of all possible reading frames is used (i.e., a protein probe vs a nucleotide database), some programs permit no control over stop codon translation and they translate all three stop codons as X, so, again, the ability to specifically identify the UGA codons of potential selenoproteins is lost.

Many of these difficulties would be eliminated if the single-letter amino acid code for Sec, U, was consistently used in databases and included in translation tables. However, that would still leave the problem of the need for a new generation of mutation matrices that include parameters for U-U matches, U-C matches, and U versus all the other amino acids. However, Sec is so rare in proteins that developing appropriate matrices (e.g., based on observed substitution probabilities in aligned blocks of protein sequences) might still be difficult to achieve.

Guided by practical considerations, we have dealt with these limitations in two ways. For database searching, we have modified existing matrices and translation tables to reflect the existence of the Sec residue, as detailed in Steps 1 and 2 below. However, for sequence

alignment and significance assessment by random shuffling, we simply treat the UGA codon as a Cys, then convert the C back to a U in the published alignment. Because Sec is so rare, and a U-U match *should* be weighted more heavily than a C-U or C-C match, this treatment of Sec as simply another Cys consistently leads to an underestimation of the significance of the alignments (see below). Particularly with Cys-rich sequences, randomly shuffled sequences will score higher than they should if a randomly placed C can line up with a conserved Sec that is treated as C by the program. Nonetheless, in the absence of a rigorously determined matrix for accurately scoring U-U versus U-C matches, and so forth, we would rather slightly *underestimate* significance scores than be accused later of *overestimating* them by use of an *ad hoc* modification of an existing matrix. However, by giving extra weight to matches to actual UGA codons, such a modified matrix can be a useful tool for locating potential selenoprotein coding regions in a database search. Guided by these considerations, we have developed the approach described below for locating gene regions with homology to GPx in viral sequence databases. A similar approach could be used to discover new homologs of other known selenoproteins in existing genomic data.

General Database Search Method for Finding Glutathione Peroxidase Modules in RNA Viruses

Step 1. In order that UGA can be distinguished from UAA or UAG stop codons in translational searches, modify the search program's translation table and define Sec (U) as a unique translation for the UGA codon. This letter (U) should also be used as the symbol for Sec in the modified mutation matrix such as PAM or Blosum (see Step 2).

Step 2. Modify the mutation matrix by creating a new row of substitution costs for Sec, represented by the letter defined in the translation table (U). Because Sec has very similar structure and characteristics to Cys, as compared with other amino acids, the value for any amino acid mutated to a U is taken to be the same as for mutation to a C, except for the case of a U-U match, in which the value is arbitrarily modified to be higher than that for the C-C or C-U match (e.g., perhaps 50% higher than the C-C value). By assigning a higher score for a U-U match in the matrix, the selectivity and sensitivity for locating selenoprotein modules in the database (Step 3) is significantly increased.

Step 3. Generate a multiple alignment of glutathione peroxidases to reveal conserved sequence motifs in the family and, with the aid of the crystal structure *(19)* and available biochemical and structural information on the catalytic mechanism *(20),* locate conserved active-site residues and regions (e.g., SLUG / SYUG region, CNQFGKQ region, etc.). These will be used to design database probes.

Step 4. Select conserved or active-site regions consisting of 10-15 amino acids, including the region spanning the Sec residue, and form a combinatorial search pattern by shifting the amino acid in each column of the multiple alignment (in essence, this involves generating a set of search probes that represent the most common active-site or sequence-motif variants in the GPx protein family).

Step 5. Conduct a newly formed search pattern search against viral public domain databases in both GenBank and EMBL, using framesearch, tfasta, and tblast search tools with a

modified translation table and mutation matrix from Step 2. Searches with a query containing Sec (U) are more likely to yield highly ranked matches to potential and actual homologs by locating genuine matches between a Sec residue in the probe and UGA codons in the targets, in desired sequence contexts.

Step 6. The top 200 hits from the database search are carefully analyzed; when many high hits belong to a specific type of virus, subsequent analysis involves (a) study of the viral genome to see if this target sequence could be expressed by any known mechanism *(6),* (b) alignment of this family of viruses, to verify that the potential gene is conserved, at least within a viral subtype or genotype, and (c) assessment of the statistical significance of the alignment between the candidate viral sequence and the GPx family (see the next subsection). If the similarity or shuffling statistic is low, repeat the cycle from Step 3 until all the combinatorial search patterns have been used.

Assessment of Statistical Significance of Sequence Similarities Between Putative Homologs and the Aligned GPx Protein Family

To assess the statistical significance of a proposed homology between a protein sequence A and a protein family B, we apply the method of Zhang et al. *(21).* This approach follows the established procedure of comparing the score of an optimal alignment between A and B to the score of an optimal alignment between a random sequence A and B, where A is constrained to have the same amino acid composition as A. Briefly, the method involves (1) the optimal alignment of the query sequence against the prealigned family, followed by (2) repeated shuffling of the query sequence and (3) calculation of a shuffling statistic to assess the significance of the query-family homology. We have found this method to be consistently more sensitive for assessing distant homology than the conventional approach based on pairwise sequence comparison *(21).*

RESULTS AND DISCUSSION

Coxsackie Virus B

Taylor et al. reported a potential GPx module encoded in the +1 reading frame in several strains of coxsackie virus B (CVB), including the CVB3 and CVB4 strains *(6).* In CVB, this sequence overlaps a region of the main polyprotein reading frame that encodes vp3, one of the picornavirus structural (capsid) proteins. The analysis also suggested the potential for expression of a GPx-vp3 fusion protein, via a -1 frameshift from the GPx coding region in the +1 reading frame, into the zero frame (see ref. 6 and Figs. 2-4 therein for details).

In both CVB3 and CVB4, downstream of the truncated GPx coding region (N-terminal active-site sequences), sequence similarities to the C-terminal region of GPx can be observed in the main reading frame (i.e., within the vp3 protein sequence itself). In both the CVB3 and CVB4 strains that have start codons in the GPx-related ORF, there are potential -1 frameshift signals (see Fig. 4 of ref. 6) that could permit this GPx module to be expressed as a fusion protein, with the GPx module fused to the C-terminal half of vp3, in which certain residues might contribute to the GPx activity because they match conserved GPx residues in the alignment (Fig. 1). Here,

we have assessed the significance of the sequence similarity between this hypothetical CVB3 GPx-vp3 fusion protein and an aligned set of plasma and cellular Se-GPx sequences.

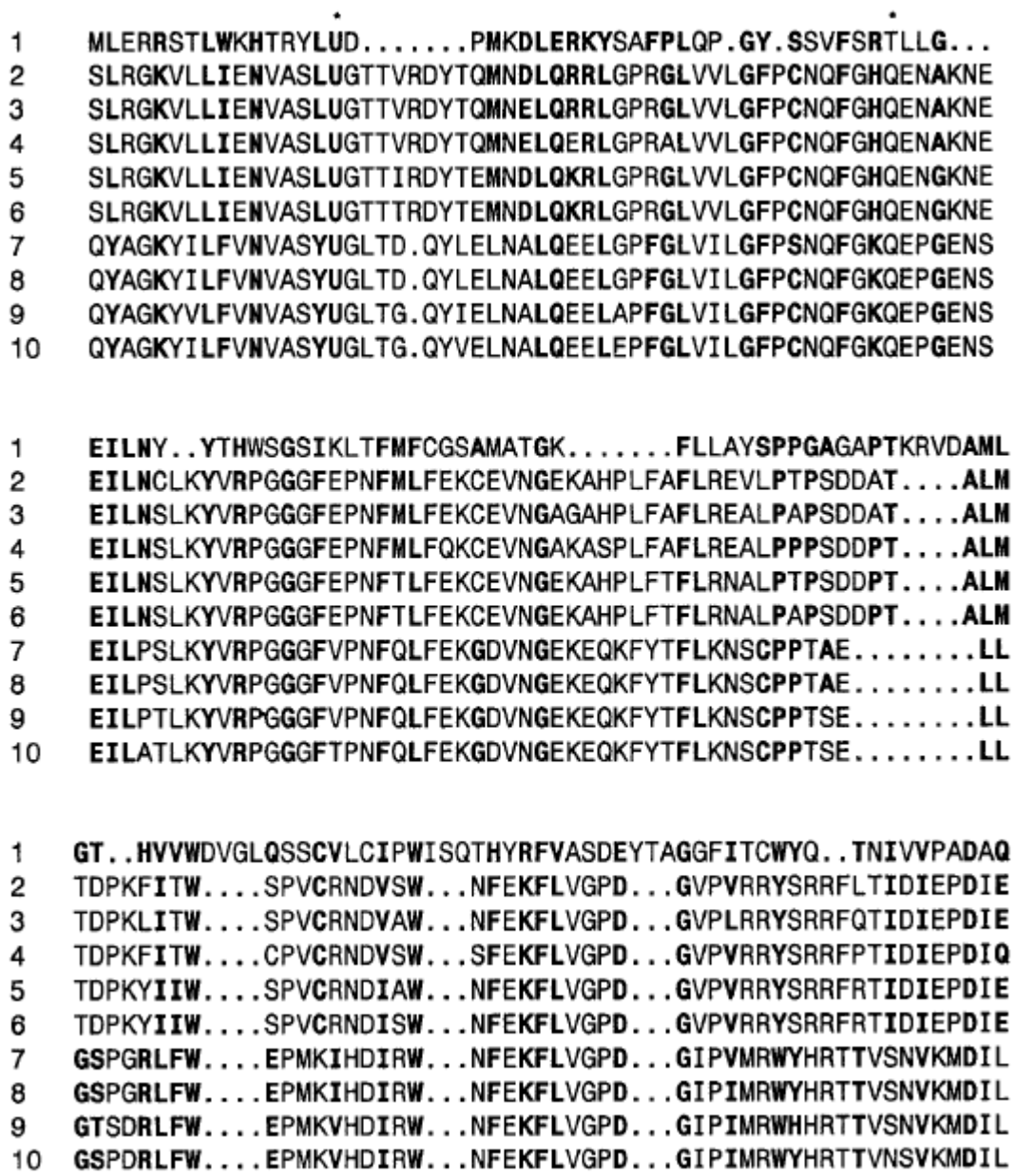

Fig. 1. Multiple sequence alignment of a hypothetical coxsackievirus B3 (CVB3) encoded GPx-vp3 fusion protein as predicted by Taylor et al. *(6),* shown here as sequence 1, versus a set of representative mammalian selenium-dependent glutathione peroxidase (GPx) sequences (#2-6 are cellular GPx, #7-10 are plasma GPx). Bold highlighting within a column indicates that a viral residue in line 1 is similar or identical to one or more of the aligned GPx sequences (this also applies to the subsequent figures). The asterisks indicate the position of the conserved catalytic Sec (U), Gln (Q), and Trp (W) residues. The Gln is not conserved in the CVB3 sequence; however, it is substituted by a Thr which has similar H-bonding ability to Gln. Alternatively, another nearby residue, such as the Asp (D) adjacent to the Sec in CVB3, may compensate for the missing Gin. The shuffling statistic (Z-score) for the complete alignment shown is 3.6 SD, relative to the average alignment score expected for random sequences of identical composition.

The computer-generated alignment is shown in Fig. 1. The shuffling statistic for the alignment is 3.6 SD (standard deviation), suggesting that there is considerable similarity between this hypothetical fusion protein and the GPx family. Notable are the conservation of several Trp residues (W) in the bottom panel (C-terminal region) of the alignment. One of these, at the position indicated by an asterisk, is a critical region-3 active site residue of GPx. This degree of similarity is perhaps remarkable, considering that 70% of this viral sequence is part of a viral structural protein, vp3, which would have to serve a dual function by also contributing structural and catalytic residues to the viral GPx activity when fused to the hypothetical N-terminal GPx module.

If this fusion protein is synthesized in vivo, a small percentage of vp3 protein molecules would contain a fused GPx module (the actual percentage being determined by frameshift efficiency and Sec availability). Because vp3 is a picornavirus structural protein present in about 60 copies per virion, this implies that GPx activity may be associated with the CVB3 virion itself. This potential GPx gene in CVB3 may also be significant in regard to the specific mutations observed by Beck et al. in the conversion of an avirulent strain into a more cardiovirulent strain *(15)* because one of the mutations consistently associated with the phenotypic change was in the vp3 gene, which might, in some way, modify the GPx-related activity in the hypothetical GPx-vp3 fusion protein, thus contributing to the greater virulence observed for the mutated strain.

HIV-1

Taylor et al. have now shown that one of the potential selenoprotein genes predicted in HIV in 1994 *(10)* is a virally encoded homolog of GPx, the prototypical mammalian selenoprotein *(18).* This gene is encoded overlapping the HIV-1 *env* gp41 coding region in the -1 reading frame (thus, the gene was tentatively named *env-fs,* for *env* frameshift) and contains a single UGA codon (potentially encoding Sec) near the middle of a highly conserved novel ORF of about 120 amino acids, with a conserved -1 frameshift signal near its 5' end. The sequence encoded in this HIV-1 gene region contains a common variant of the GPx active-site consensus sequence, characteristic of the plasma or extracellular type of GPx, spanning the catalytic Sec (see alignment in Fig. 2; active-site sequence YUGAT in HIV-1 vs YUGLT in GPx, where U -- Sec). The region of highest similarity, corresponding to the GPx active site, occurs precisely around the location of the single conserved UGA codon in the -1 reading frame of the gp41 coding region of HIV-1. The UGA codon in *env-fs,* corresponding to the catalytic Sec of GPx, is highly conserved in the predominant North American-European HIV-1 strains (subtype B), but other HIV-1 subtypes and most HIV-2 subtypes have a conserved arginine codon (AGA) at this position This substitution is also consistent with peroxidase activity, because arginine is the active-site residue of many peroxidases, such as horseradish peroxidase.

Using a sensitive multiple-sequence comparison method *(21),* a strong similarity between the entire HIV-encoded *env-fs* sequence and the aligned set of GPx sequences was demonstrated. As aligned previously by Taylor et al. *(18),* the similarity score of this novel HIV sequence versus an aligned group of GPx sequences is 5 SD above the average similarity score of randomized sequences of identical composition.

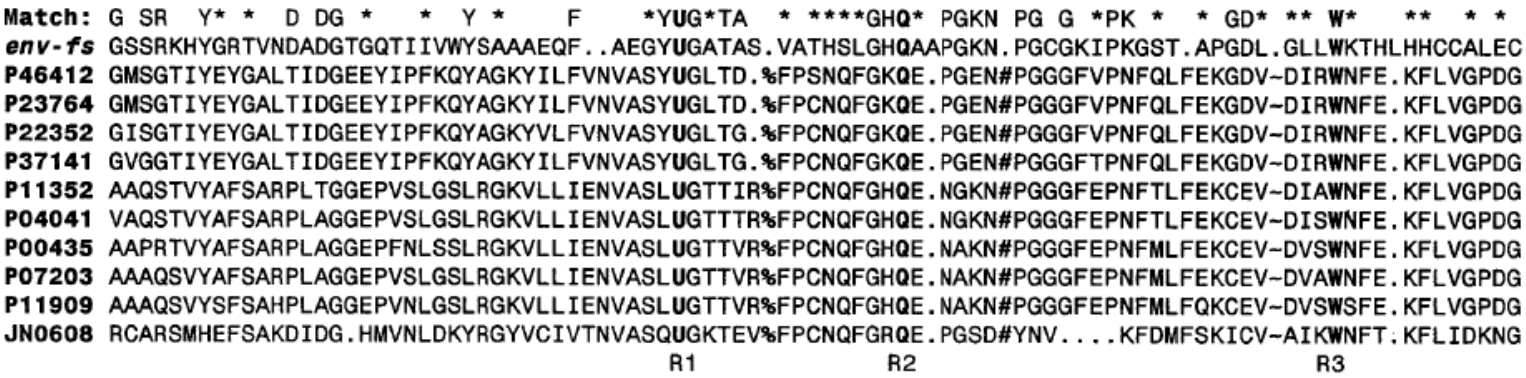

Fig. 2. Sequence alignment of hypothetical HTV-1 *env-fs* sequence versus various selenium-dependent glutathione peroxidases. The HIV-1 sequence shown as *env-fs* is as given in Table 1 of Taylor et *al. (10),* beginning right after a predicted potential protease cleavage site. It is aligned with a set of eukaryotic glutathione peroxidases (GPx) from various sources (single-letter amino acid code, with U as the symbol for selenocysteine, Sec, encoded by the UGA codon in RNA). *env-fs* amino acids identical to one or more of the aligned GPx sequences are shown as letters in the "Match" line above the alignment; similar residues are indicated by an asterisk. Note the matches in the region of the conserved catalytic selenocysteine, with *env-fs* residues YUGAT, compared to YUGlT in the plasma GPx sequences (note that, here, UGA means Sec–Cly–Ala, not the UGA codon). The most common variant in the predominant North American-European HIV-1 subtype B is SYUGAT instead of GYUGAT, an even better match to many of the GPx (active-site motif usually SLUGTT or SYUGLT). The N-terminal of mature GPx protein begins at the fourth residue in the alignment. The three active-site regions are shown below the alignment, labeled Rl- R3. All three regions and their essential conserved active-site amino acids, Sec, Gln, and Trp, are represented in the truncated *env-fs* sequence (the active-site residues, U, Q, and W respectively, are shown highlighted in bold). There are three large deletions in *env-fs* relative to the GPx sequences, at the locations indicated by the symbols %, #, and ~ in the alignment. These deletions in *env-fs,* involving 19, 11, and 41 residues in %, #, and ~, respectively, lie between the conserved active-site regions (YUG, GHQ, and W) and are distant from the active site in three-dimensional space. Based on an examination of the bovine GPx crystal structure, all three deletions are structurally feasible, although the third deletion at ~ might lead to loss of the ability to form a tetrameric enzyme (this would not be unprecedented because phospholipid hydroperoxide GPx is known to function as a monomer). As computed using the alignment shown (i.e., omitting the GPx regions at %, #, and ~, where there are large gaps in the truncated HIV-1 GPx homolog), the total similarity score of the *env-fs* sequence to the aligned GPx sequences is 6.3 SD above the average similarity score of 100 randomly shuffled sequences of identical composition, each optimally aligned by the same algorithm (see text).

For the current study, we have examined alternative alignments of the C-terminal region of *env-fs* in the light of the bovine GPx crystal structure *(19)* and identified the viral sequence matching a third GPx active-site region with a conserved Trp (W), shown as GPx region 3 (R3) in the alignment of Fig. 3, which reveals a total of three significant internal deletions in the HIV-1 sequence relative to GPx. These deletions in *env-fs,* involving 19, 11, and 41 residues, respectively, lie between the three conserved active-site regions (YUG, GHQ, and W) and project away from the active site in three-dimensional space. All three deletions are structurally feasible, although the third and largest deletion just upstream of region 3 would involve loss of a domain that participates in the formation of the tetrameric GPx enzyme. This suggests that the HIV-1 GPx homolog probably does not function as a tetramer, which would make it similar to the phospholipid hydroperoxide GPx, which is known to function as a monomer *(1)* and which also has a deletion in this same region, although not as large as the deletion in *env-fs.*

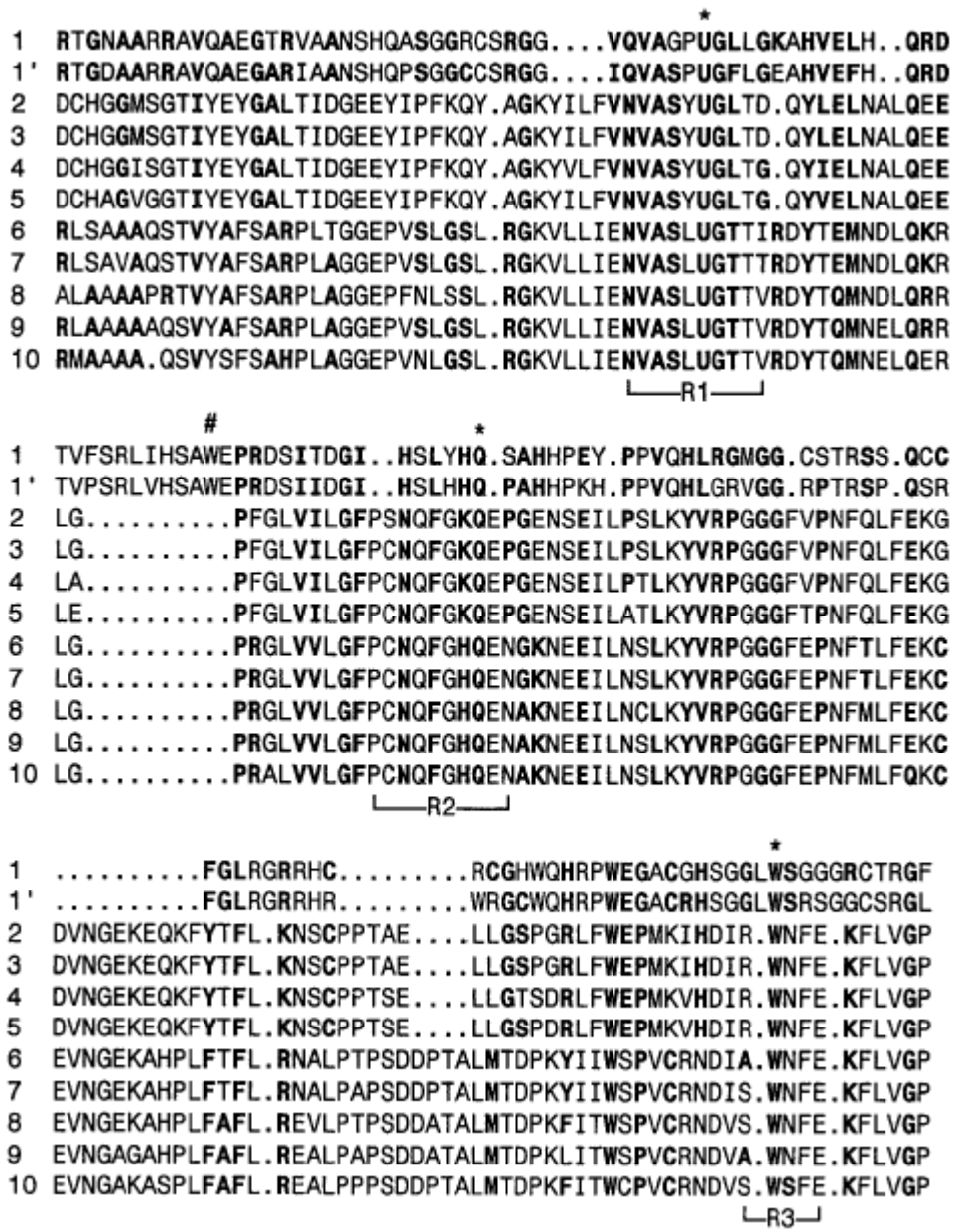

Fig. 3. Multiple sequence alignment of a hypothetical hepatitis C virus (HCV) encoded GPx homolog, with two common genotype lb variants shown as sequences 1 and l', versus a set of Se-GPx sequences (#2-5 are plasma GPx, #6-10 are cellular GPx). The asterisks i.ndicate the positions of the conserved catalytic Sec (U), Gin (Q}, and Trp (W) residues in active-site regions Rl-R3. The Trp residue at the position indicated by # is well conserved in HCV sequences and, in the many viral isolates that are truncated al the C-terminal, may take over the function of the active-site Trp in R3.

Using the alignment shown in Fig. 2 (i.e., omitting the GPx regions for which there are large gaps in the truncated HIV-1 GPx homolog), the total similarity score of the *env-fs* sequence to the aligned GPx sequences is 6.3 SD above the average similarity score of randomly shuffled sequences of identical composition. This suggest that the *env-fs-GPx* similarity is even higher than the 5 SD estimated previously by Taylor et al. *(18).* The alignment shown, with a Trp residue (W) in the HIV-1 sequence aligning with the region-3 active-site Trp of GPx, is strongly supported by the fact that the HIV-1 Trp residue in question is highly conserved in the main

group of HIV-1 sequences, being 100% conserved in HIV-1 subtype B. This conservation also provides compelling theoretical evidence that this is a functional gene, because the UGG Trp codon in the -1 frame is in the context of the sequence UCUGGA, which encodes Ser-Gly in the HIV envelope protein. Because of the degeneracy of the genetic code, Ser could be encoded using any base in the third codon position (i.e., UCN), so conservation of a U in over 40 HIV-1 subtype B sequences can only be explained by a functional requirement, such as the need for the UGG in the -1 frame. Because this region is downstream of the *rev-responsive* element RNA structure and because the GPx homology region is the *only* other known or potential coding region overlapping this region of the *env* gene, this conservation strongly suggests that the GPx gene is functional and that the upstream UGA must be translated as an amino acid; otherwise, the Trp-containing region could not be expressed and there would be no basis for evolutionary selection to conserve the Trp codon in the -1 frame.

This hypothetical HIV-1 GPx gene has now been cloned for functional studies in transfected cells. When the encoded sequence is expressed as a selenoprotein in mammalian cells, highly significant increases in measured GPx activity are observed *(22).* That observation is in conflict with the conclusions of a recent study of selenoprotein expression in HIV-1 infected and uninfected T cells by Gladyshev et al., which purported to find no evidence of HIV-encoded selenoproteins *(23).* Those authors did note that in HIV-infected cells, there was a decline in levels of cellular selenoproteins and an increase in "low-molecular-mass" Se compounds. The latter observation is of particular interest because several potential HIV selenoproteins are predicted to be of low molecular mass, from 7 to 9 kDa, including the 9-kDa isoform of the HIV GPx homolog corresponding to the sequence shown in Fig. 1 (i.e., active in functional assays) *(22).* Thus, we vigorously dispute not the data but the conclusions of Gladyshev et al., as detailed in a critique of their study *(24).* However, their demonstration that levels of cellular selenoproteins decline in HIV-infected cells is actually consistent with the predictions of the HIV selenoprotein theory and could help to explain why the effects of HIV infection may be exacerbated in individuals who are Se deficient.

Hepatitis C Virus

We report here for the first time a GPx gene in hepatitis C virus (HCV), which is a very common infection in the United States (about 1.5% or 4 million people are seropositive). As in the case of HIV-1, in HCV the GPx gene is in the -1 reading frame overlapping a known gene (the NS4b gene in the case of HCV), contains an in-frame "stop" codon, UGA, that can also encode selenocysteine, and lacks an apparent start codon, thus explaining why these genes have escaped detection up to now. The Se-dependent GPx sequence and UGA codon are highly conserved in HCV genotype lb, which is predominant in North America.

The putative HCV GPx sequence has regions of very high similarity to known GPx sequences; the similarity encompasses the entire sequence (Fig. 3) but is strongest in the active-site regions, for which the similarity is statistically significant at 6.2 SD relative to random sequences of similar composition or 6.7 SD if compared only to the mammalian plasma GPx enzymes. The HCV GPx (active-site region-1 amino acid sequence VQVASPUGLLG) is perhaps most similar to the human plasma GPx (active-site R1 sequence VNVASYUGLTG, where U signifies the Sec codon), although it has features of both the plasma and cellular GPx types.

Unlike the HIV-1 example, where there is a potential -1 frameshift sequence for expression of the GPx module *(10,18),* it seems likely that the HCV GPx is expressed via RNA editing analogous to that involved in the expression of paramyxovirus P genes *(25-27).* This type of RNA editing, involving a "stuttering" mechanism of the viral polymerase *(25),* can insert one or more extra nucleotides in a subset of RNA transcripts, thus bringing an overlapping-reading frame in-frame to the main polyprotein. We have identified a potential RNA editing site in HCV near the beginning of the GPx ORE consisting of a heptameric homopyrimidine sequence with a triplet of C bases, similar to the paramyxovirus RNA editing consensus sequence proposed by Vidal et al. *(25).* There is also a potential RNA pseudoknot in HCV, about 15 bases upstream of the homopyrimidine sequence in the plus RNA strand, which could induce a pause during minus strand synthesis, facilitating the melting and offset reannealing of the template-primer, in a manner similar to the role of pseudoknots in frameshifting during protein translation. By this mechanism, edited minus strand RNAs formed by this editing process would become templates for synthesis of mRNAs from which the GPx module could be translated, rather than the NS4b gene product and the downstream proteins of the unedited transcripts.

Finally, this proposed RNA editing site is just a few residues downstream of a known HCV protease cleavage site, at the NS4a-NS4b junction. Thus, when translated from an edited RNA, the HCV GPx would be cleaved from the viral polyprotein only a few residues upstream from the beginning of the region shown in Fig. 3. This proteolytic, processing to form an independent protein suggests a predominantly intracellular and possibly intravirion role for the HCV GPx.

Measles Virus

A database search with the cellular "SLUG" variant of the GPx active-site R1 motif gave isolates of the measles virus fusion (F) protein gene as highly ranked matches, with the 3' terminal UGA stop codon of the fusion protein aligning with the essential Sec of GPx. As shown in Fig. 4, the similarity encompasses the peptide encoded downstream of the UGA and is statistically significant at 5.0 SD. Examination of five morbillivirus F gene sequences that we were able to locate in GenBank showed that the UGA codon is conserved in all cases (listed by Gen-Bank accession number: M14915, $58435, Z66517, U62652, and DOOO90). Four of the five have a second downstream in-frame UGA that aligns with the conserved Cys residue in GPx active-site region 2 (CNQ motif; Fig. 4). Although these putative measles-virus GPx-related sequences are truncated at the C-terminal, lacking the active-site region 3, they would be extended at the N-terminal by fusion to the F protein, which could provide supporting structure and catalytic residues in a manner analogous to that discussed earlier for the CVB GPx-vp3 fusion protein. If this proves to be a functional GPx module, it would be an unique GPx variant if the second UGA also encodes Sec, because all known GPx enzymes have at most one Sec residue.

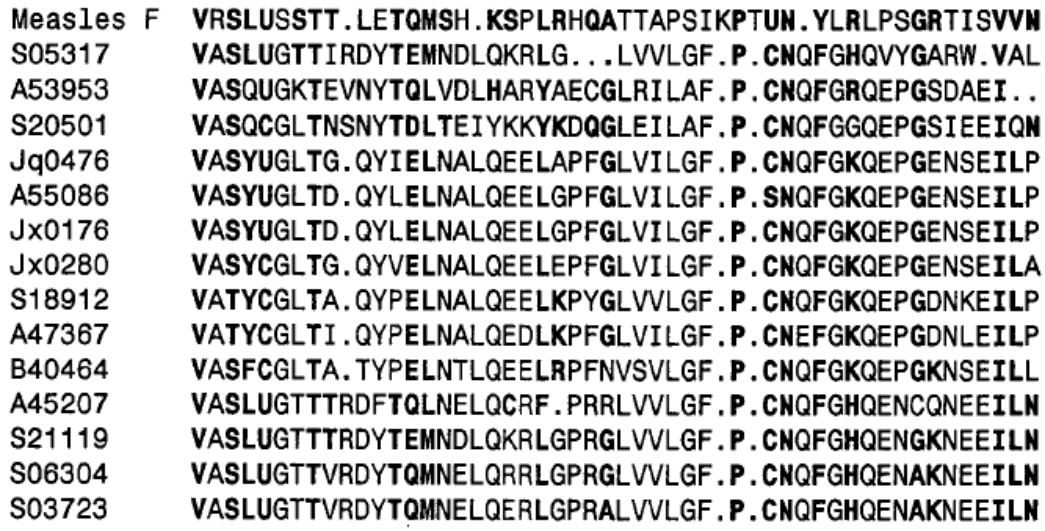

Fig. 4. Sequence alignment for hypothetical measles-virus selenoprotein with various glutathione peroxidases. Alignment score = –133; average alignment score for 100 randomly shuffled measles sequences = 391 ± 105; thus, shuffling statistic (z-score) = 5.0 SD. Parameters used are blosum68 matrix, gap startup penalty 6, and gap extension penalty 2.

HIV-2

Another significant match to a variant of the GPx active-site R1 "SLUG" motif was found in HIV-2 isolates, at the 3' end of the *env* gene. This is of considerable interest because, unlike HIV-1, HIV-2 does not encode a Se-GPx in the overlapping -1 reading frame of the *env* gene; thus, both HIV-1 and HIV-2 may be accomplishing a similar effect (i.e., encoding an envelope-associated Se-GPx module) by slightly different means. As observed for the measles-virus membrane fusion (F) protein gene, in the case of the HIV-2 *env* gene, the 3' terminal UGA stop codon of the known protein aligns with the essential Sec of GPx. As shown in Fig. 5, the similarity encompasses the peptide encoded downstream of the UGA and is statistically significant at 4.5 SD. Examination of multiple alignments of all known HIV-2 envelope genes and the consensus sequences of the various HIV-2 subtypes showed that the UGA codon is totally conserved in all HIV-2 subtypes (Fig. 6). As discussed earlier, for the similar situation in measles virus, the putative HIV-2 *env* GPx module is truncated at the C terminal, lacking the active-site region 3, but would be extended at the N-terminal by fusion to (in this case) the *env* transmembrane protein, which could provide supporting structure and catalytic residues to this truncated module. This encoding of the putative HIV-2 GPx module, fused to the C-terminus of the intracellular domain of the *env* protein, suggests that, if expressed, this module would be located just inside the membrane of infected cells and viral particles, attached to the envelope protein. This would be analogous to but distinct from the HIV-1 GPx module, which would be expressed as an *env* gp 120 fusion protein and would more likely be located on the outside of cells and virions. Appropriately, the HIV-2 GPx is closer to the cellular GPx, whereas the HIV-1 sequence is closer to the plasma GPx, consistent with this cellular versus extracellular difference in predicted location.

Fig. 5. Sequence alignment of hypothetical HIV2 *env* UGA read-through consensus sequence for subtype A with various glutathione peroxidases. Alignment score = 309; average alignment score for 100 randomly shuffled HIV-2 sequences = 716 ± 90; thus, shuffling statistic (z-score) = 4.5 SD. Parameters used are blosum68 matrix, gap startup penalty 7, and gap extension penalty 2.

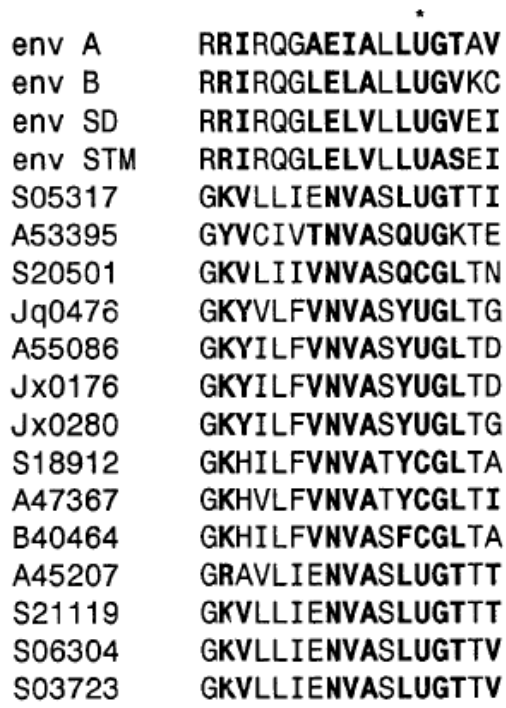

Fig. 6. Sequence alignment of potential HIV-2 *env* C-terminal UGA read-through motifs from four consensus sequences of different .HIV-2 subtypes (A, B, etc.) versus various Sc-dependent and Se-independent glutathione peroxidase sequences. U = Sec.

SIGNIFICANCE AND CONCLUSIONS

By structurally guided comparative sequence analysis, we have provided compelling evidence in support of the hypothesis that Se-dependent GPx modules are encoded in a number of RNA viruses, including potentially serious pathogens like HIV-1 and the hepatitis C virus. We have

focused our study on examples where sequences of multiple viral isolates are available, so that conservation of the putative GPx-related features, at least within viral subtypes or genotypes, can be verified. A weakness of the current study is that we have not attempted to identify specific viral mechanisms or RNA structure features that might be involved in recoding the UGA codon for translation as Sec. In part, this reflects a general lack of knowledge and lack of precedents for such mechanisms in RNA viruses. We believe that, given the significant differences in Sec translation mechanisms between bacteria and eukaryotes and the knack that RNA viruses have for streamlining various coding functions because of their limited genome size, these viruses may ultimately prove to have somewhat unique mechanisms for Sec insertion, which may be difficult to identify by a purely theoretical analysis. The current predictions of GPx coding potential in RNA viruses are primarily based on the presence and conservation of significant protein sequence homology, combined with evidence that the gene regions in question may be expressed by known mechanisms such as frameshifting, RNA editing, and read-through suppression. The specific mechanisms by which the UGA codon may be decoded as Sec during the expression of these putative viral GPx modules remain to be elucidated by future studies.

There are several possible benefits that a virus might gain by encoding a GPx. A virally encoded GPx could help a virus defend against free-radical-mediated attacks on infected cells by the immune system *(4),* and, 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 enveloped viruses, which are more susceptible to membrane lipid peroxidation once they have budded off the host cell and have lost the benefit of cellular antioxidant defenses.

In regard to the latter possibility, it is significant that in several cases we have identified here, the GPx module is associated with an outer-coat protein or cell-fusion-related protein of the virus (i.e., the vp3 capsid protein of coxsackievirus B3, the envelope genes of HIV-1 and HIV-2, and the fusion protein of measles virus). This suggests that the GPx is useful either (1) to protect the virus from reactive oxygen species in the extracellular environment, or (2) during attachment to target cells (which in some cases are known to release a "respiratory burst" of peroxide when contacted by viruses), or (3) for both of the above. Such extracellular roles would also be consistent with the slightly greater similarity between several of the viral GPx enzymes and the plasma versus cellular type of GPx. One problem that has concerned investigators is the fact that glutathione (GSH) levels are very low in blood, so it has been unclear how plasma GPx can reduce extracellular peroxides without sufficient GSH available as a reducing agent. However, new data indicate that plasma GPx can use glutaredoxin and thioredoxin for thiol-reducing equivalents *(28)* and, thus, function efficiently under conditions of low GSH.

A viral GPx might also have a repressive effect on viral replication, because it is known that oxidative stress (e.g., H202 exposure) activates the replication of HIV-1 and other viruses: A viral GPx would reduce oxidant tone, thus reducing viral activation. There is a firm experimental basis for this possibility in regard to HIV-1 because in latently infected cell cultures, it has been shown that increasing GPx expression by Se supplementation inhibits viral cytotoxic effects and the reactivation of HIV-1 by hydrogen peroxide, decreases activation of the transcription factor NF-κB (an important cellular transactivator of HIV-1), and protects against activation of HIV-1 by tumor necrosis factor (TNF) *(29).* Similar results showing selenite inhibition of TNF-activated HIV expression in chronically infected T cells and monocyte-derived macrophages

have more recently been reported by Hatfield and coworkers *(30).* Thus, HIV-1 could use its own GPx gene to negatively regulate its transcription by the same mechanism, which could be contributing to the above observations. This might permit the virus to maintain a low profile during times of immune activation (associated with high Se status), as well as to fend off free-radical-mediated apoptotic attacks *(4).* In contrast to the results observed in the chronically infected cell model, where exogenous; selenite leads to increased cellular GPx and the inhibition of HIV-1 proviral transcription *(29,30),* Sandstrom et al. *(31)* reported that overexpression of a cellular GPx gene in a T-cell line leads to accelerated HIV-1 replication and emergence of cytopathic effects. Thus, an HIV-encoded GPx gene, particularly if envelope associated, might assist the virus in the acute phase of cellular infection by a similar mechanism.

The existence of a sequence in HIV-1 with highly significant similarity to a known selenoprotein, GPx, along with the extensive clinical data strongly linking HIV-1 disease progression to the Se status of the host (as reviewed in refs. 6 and *18),* provide compelling support for the hypothesis that there is a unique interaction between Se and HIV and that Se may be an unusually critical nutritional factor in HIV disease progression *(32).* Yet, despite the certainty that the virus would only encode a selenoprotein if it conferred an evolutionary advantage, it is fascinating that all the clinical data show that it is Se deficiency that correlates with a negative outcome in HIV-1 infection *(33-35).*

In regard to hepatitis C, the existence of a virally encoded Se-GPx, well conserved in genotype lb, may help to explain certain clinical aspects of HCV infection. Significantly, genotype lb is associated with the highest risk of progression to cirrhosis and hepatocellular carcinoma, and poor response to interferon. An HCV-encoded GPx gene may help explain why oxidant stressors like alcoholism and iron overload are associated with HCV disease progression. The best direct evidence consistent with an HCV-Se link are the clinical data of Look et al., who found that in HIV+ patients, the progressive decline in Se levels characteristic of HIV infection was *greater in those with HCV coinfection,* who "showed markedly lower Se concentrations compared to those without concomitant HCV-infection" *(34).*

The *M. contagiosum* precedent for a virally encoded selenoprotein means that the possibility of additional coding potential in HIV, HCV, and other pathogenic human viruses, involving suppression of certain highly conserved UGA codons, deserves serious consideration. If active, as appears highly likely in the case of HIV-1 *(22),* such genes would establish the basis of an unprecedented role for Se in the biochemistry and regulation of chronic viral infections like HIV and HCV. Our identification of a potential GPx gene in the putative Keshan disease cofactor CVB3 (Fig. 1) provides a new basis for interpreting the results of Beck and co-workers cited earlier, showing that Se deficiency increases the virulence of CVB3 infection *(14,15).* Because Keshan disease was thought to be simply a Se-deficiency disease, but proved to have a viral cofactor, it would not be unprecedented if AIDS and viral hepatitis ultimately prove to be viral diseases with Se deficiency as a major cofactor.

REFERENCES

1. R. A. Sunde, Selenium, in *Handbook of Nutritionally Essential Minerals,* B. L. O'Dell and R. A. Sunde, eds., Marcel Dekker, New York, pp. 493-556 (1997).

2. R. J. Turner and J. M. Finch, Selenium and the immune response, *Proc. Nutr. Soc.* 50, 275-285 (1991).

3. E. W. Taylor, Selenium and viral diseases: facts and hypotheses, *J. Orthomol. Med.* 12, 227-239 (1997).

4. T. M. Buttke and P. A. Sandstrom, Redox regulation of programmed cell death in lymphocytes, *Free Radical Res.* 22, 389-397 (1994).

5. M. Roy, L. Kiremidjian-Schumacher, H. I. Wishe, M. W. Cohen, and G. Stotzky, Supplementation with selenium and human immune cell functions. I. Effect on lymphocyte proliferation and interleukin 2 receptor expression. *Biol. Trace Element Res.* 41, 103-114 (1994).

6. E. W. Taylor, R. G. Nadimpalli, and C. S. Ramanathan, Genomic structures of viral agents in relation to the biosynthesis of selenoproteins, *Biol. Trace Element Res.* 56, 63-91 (1997).

7. M. J. Berry, G. W. Martin III, and S. C. Low, RNA and protein requirements for eukaryotic selenoprotein synthesis, *Biomed. Environ. Sci.* 10, 182-189 (1997).

8. M. J. Berry and P. R. Larsen, Recognition of UGA as a selenocysteine codon in eukaryotes: a review of recent progress, *Biochem. Soc. Trans.* 21, 827-832 (1993).

9. Bock, K. Forchhammer, J. Heider, W. Leinfelder, G. Sawers, B. Veprek, et al., Selenocysteine: the 21st amino acid, *Mol. Microbiol.* 5, 515-520 (1991).

10. E. W. Taylor, C. S. Ramanathan, R. K. Jalluri, and R. G. Nadimpalli, 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.* 37, 2637-2654 (1994).

11. E. W. Taylor, C. S. Ramanathan, and R. G. Nadimpalli, A general method for predicting new genes in nucleic acid sequences: application to the human immunodeficiency virus, in *Computational Medicine, Public Health and Biotechnology,* Ser. Math. Biolo. Med. 5, 285-309, M. Witten, ed., World Scientific, Singapore, (1996).

12. S. Ramanathan and E. W. Taylor, Computational genomic analysis of hemorrhagic fever viruses: viral selenoproteins as a potential factor in pathogenesis, *Biol. Trace Element Res.* 56, 93-106 (1997).

13. J. Bai, S. Wu, K. Ge, X. Deng, and C. Su, The combined effect of selenium deficiency and viral infection on the myocardium of mice, *Acta Acad. Med. Sin.* 2, 29-31 (1980).

14. M. A. Beck, P. C. Kolbeck, L. H. Rohr, Q. Shi, V. C. Morris, and O. A. Levander, Benign human enterovirus becomes virulent in selenium-deficient mice, *J. Med. Virol.* 43, 166-t70 (1994).

15. M. A. Beck, Q. Shi, V. C. Morris, and O. A. Levander, Rapid genomic evoluton of a non-virulent Coxsackievirus B3 in selenium-deficient mice results in selection of identical virulent isolates, *Nature Med.* 1, 433-436 (1995).

16. T. G. Senkevich, J. J. Bugert, J. R. Sisler, E. V. Koonin, G. Darai, and B. Moss, Genome sequence of a human tumorigenic poxvirus: prediction of specific host response-evasion genes, *Science* 273, 813-816 (1996).

17. J. L. Shisler, T. G. Senkevich, M. J. Berry, and B. Moss, Ultraviolet-induced cell death blocked by a selenoprotein from a human dermatotropic poxvirus, *Science* 279, 102-105 (1998).

18. E. W. Taylor, A. Bhat, R. G. Nadimpalii, W. Zhang, and J. D. Kececioglu, HIV-1 encodes a sequence overlapping env gp41 with highly significant similarity to selenium-dependent glutathione peroxidases, *J. AIDS Hum. Retrovirol.* 15, 393-394 (1997).

19. O. Epp, R. Ladenstein, and A. Wendel, The refined structure of the selenoenzyme glutathione peroxidase at 0.2 nm resolution, *Eur. J. Biochem.* 133, 51-69 (1983).

20. K. D. Aumann, N. Bedorf, R. Brigelius-Flohe, D. Schomburg, and L. Flohe, Glutathione peroxidase revisited--simulation of the catalytic cycle by computer-assisted molecular modelling, *Biomed. Environ. Sci.* 10, 136-155 (1997).

21. W. Zhang, J. D. Kececioglu, and E. W. Taylor, Assessing distant homology between an aligned protein family and a proposed member through accurate sequence alignment, *J. Mol. Evol.,* submitted.

22. L. Zhao, J. A. Ruzicka, A. G. Cox, and E. W. Taylor, HIV-1 encodes a sequence with functional glutathione peroxidase activity: implications for the link between selenium deficiency and AIDS, presented at the *4th Dresden Selenium Symposium,* Dresden, Germany, May 15-16, 1999.

23. V. N. Gladyshev, T. C. Stadtman, D. L. Hatfield, and K.-T. Jeang, Levels of major selenoproteins in T cells decrease during HIV infection and low molecular mass selenium compounds increase, *Proc. Natl. Acad. Sci. USA* 96, 835-839 (1999).

24. E. W. Taylor, Defense of the HIV selenoprotein theory: a critique of Gladyshev et al. (1999) Posted at http://bioinfo.chem.uga.edu/homepage/wtaylor.

25. S. Vidal, J. Curran, and D. Kolakofsky, A stuttering model for paramyxovirus P mRNA editing, *EMBO J.* 9, 2017-2022 (1990).

26. S. M. H. Horikamin and S. A. Moyer, Synthesis of leader RNA and editing of the P mRNA during transcription by purified measles virus, *J. Virol.* 65, 5342-5347 (1991).

27. M. Steward, I. B. Vipond, N. S. Millar, and P. T. Emmerson, RNA editing in Newcastle disease virus, *J. Gen. Virol.* 74, 2539-2547 (1993).

28. M. Bjornstedt, S. Kumar, L. Bjorkhem, G. Spyrou, and A. Holmgren, Selenium and the thioredoxin and glutaredoxin systems, *Biomed. Environ. Sci.* 10, 271-279 (1997).

29. Sappey, S. Legrand-Poels, M. Best-Belpomme, A. Favier, B. Rentier, and J. Piette, Stimulation of glutathione peroxidase activity decreases HIV Type 1 activation after oxidative stress, *AIDS Res. Hum. Retrovir.* 10, 1451-1461 (1994).

30. K. Hori, D. Hatfield, F. Maldarelli, B. J. Lee, and K. A. Clouse, Selenium supplementation suppresses tumor necrosis factor c~-induced human immunodeficiency virus type 1 replication in vitro, *AIDS Res. Hum. Retrovir.* 13, 1325-1332 (1997).

31. P. A. Sandstrom, J. Murray, T. M. Folks, and A. M. Diamond, Antioxidant defenses influence HIV-1 replication and associated cytopathic effects, *Free Radical Biol. Med.* 24, 1485-1491 (1998).

32. W. Taylor, A. G. Cox, L. Zhao, J. A. Ruzicka, A. Bhat, W. Zhang, et al., Nutrition, HIV and drug abuse: the molecular basis of a unique rote for selenium, *J. AIDS Human Retrovir.,* in press.

33. J. Constans, J. L. Pellegrin, C. Sergeant, M. Simonoff, I. Pellegrin, H. Fleury, et al., Serum selenium predicts outcome in HIV infection, *J. AIDS* 10, 392 (1995).

34. M. P. Look, J. K. Rockstroh, G. S. Rao, K. A. Kreuzer, S. Barton, H. Lemoch, et al., Serum selenium, plasma glutathione (GSH) and erythrocyte glutathione peroxidase (GSH-Px)-levels in asymptomatic versus symptomatic human immunodeficiency virus-1 (HIV-1)-infection, *Eur. J. Clin. Nutr.* 51, 266-272 (1997).

35. M. K. Baum, G. Shor-Posner, S. Lai, G. Zhang, H. Lai, M. A. Fletcher, et al., High risk of mortality in HIV infection is associated with selenium deficiency, *J. AIDS Hum. Retrovir.* 15, 370-374 (1997).